第 1 章

引　　言

数学规划是最优化理论的一个分支．在数学规划中，对 n 个实变量 $x_1, \cdots, x_n$ 的一个单值目标函数 f 进行极小化(或极大化)，这些变量可能受限制于有限个写成不等式或方程的约束．一般，我们定义一个极小化的数学规划(Mathematical Program)为

(MP)
$$\min f(x) \tag{1.1}$$

受限制于

$$g_i(x)\geqslant 0, \quad i=1, \cdots, m, \tag{1.2}$$

$$h_j(x)=0, \quad j=1, \cdots, p, \tag{1.3}$$

其中 x 是分量为 $x_1, \cdots, x_n$ 的列向量．换句话说，(MP)是求一个满足(1.2)和(1.3)的向量 x^*，使 $f(x)$ 取极小值——也就是最优值．如果出现在(MP)中的函数有一个或多个是 x 的非线性函数，我们称之为**非线性规划**．这与线性规划相反，那里的一切函数都必须是线性的．非线性规划某些基本方面的研究是本书的主题．

非线性规划问题，出现在如工程、经济、商业管理、物理科学和数学这样一些不同的学科中，或者出现在具有下述性质的其他领域中，那里需要在某些复杂（或相冲突）的情况下作出(在广义下的)决策，而这些情况又能用数学模型表示．为了说明某些类型的非线性规划，下面介绍几个例子．

非线性曲线拟合

在某些科学研究中，例如在生物学或物理学中，假设某种现象 f 作为时间的函数在实验室中被测量，还假设这现象的数学模型已经给定，并且根据模型知道 f 之值随时间 t 的变化为

$$f(t)=x_1+x_2\exp(-x_3t). \tag{1.4}$$

实验室试验的目的是根据在时刻 $t^1, t^2, \cdots, t^M$ 测得的 f 之值来求未知参数 x_1, x_2 和 x_3．这一决策过程包含要指定参数值，于是自

然地要求 x_1, x_2 和 x_3 的数值在某种意义下为最优. 例如,我们可以在最小二乘方意义下求参数的最优值,这就是说,求那些数值,使实验曲线与理论曲线的偏差的平方和为最小. 形式上,我们有非线性规划

$$\min F(x_1, x_2, x_3) = \sum_{i=1}^{M} [f(t^i) - x_1 - x_2 \exp(-x_3 t^i)]^2. \tag{1.5}$$

注意到这是一个无约束规划,解出的参数值或许是不能接受的.为了避免这一情况,可以加上一些约束形式的限制.例如,参数 x_3 限定为非负,即

$$x_3 \geqslant 0. \tag{1.6}$$

对考察的特殊现象,还可假设所提出的数学模型,只在参数选为能使 $t=0$ 时 $f(0)=1$, 才是可以接受的. 因此必须加上约束

$$x_1 + x_2 = 1, \tag{1.7}$$

受限制于(1.6)和(1.7)而求解(1.5),则是一个约束非线性规划问题,具有非线性的目标函数及线性的不等式和等式约束.

定位问题

现在假设要选定一个供应中心的位置,这中心向城市中有固定空间位置的 m 个用户提供服务. 中心供应的商品可以是电、水、牛奶或其他货物. 供应中心的定位准则是使从中心到用户的某个距离函数最小. 例如,很可能我们感兴趣的是使中心到任何一个用户的**最大距离**为最小. 由于在这个城市里货物必须沿垂直的路线(如街道)供应,合适的距离函数是所谓**矩形距离**. 作为数学模型列出来,设 (x_1, x_2) 表示供应中心的待定位置(坐标),而 (a^i, b^i) 是第 i 个用户的给定位置. 我们的问题则是

$$\min_{x_1, x_2} \{\max_{1 \leqslant i \leqslant m} [|a^i - x_1| + |b^i - x_2|]\}, \tag{1.8}$$

这个式子意味着,首先对 (x_1, x_2) 的每个可能值,必须求指标 i 使方括弧中的矩形距离最大,其次在依赖于 (x_1, x_2) 的所有最大距离中求出最小的. 另外,如果每一位置 (x_1, x_2) 都可以接受,那末问题是无约束的. 但是,有时为了简化某些表示式,以添加额外变量

和约束为代价是有利的．例如，定义一个新变量 x_0：

$$x_0=\max_{1\le i\le m}\left[|a^i-x_1|+|b^i-x_2|\right] \tag{1.9}$$

或

$$x_0\geqslant|a^i-x_1|+|b^i-x_2|,\quad i=1,\ \cdots,\ m, \tag{1.10}$$

我们就得到一个有三个变量 x_0, x_1, x_2 的非线性规划：

$$\min f(x)=x_0 \tag{1.11}$$

受限制于

$$g_i(x)=x_0-|a^i-x_1|-|b^i-x_2|\geqslant 0,\quad i=1,\ \cdots,\ m. \tag{1.12}$$

读者容易验证，问题(1.8)和问题(1.11)至(1.12)在以下意义下等价：$(x_1^*,\ x_2^*)$为(1.8)的最优解的充要条件是$(x_0^*,\ x_1^*,\ x_2^*)$为(1.11)和(1.12)的最优解，而且对某个 $k(1\leqslant k\leqslant m)$ 成立

$$x_0^*=|a^k-x_1^*|+|b^k-x_2^*|. \tag{1.13}$$

读者还能证明，通过引进更多的变量，(1.8)可以化为一个线性规划问题．

过程设计

在连续搅拌的槽式(回收混合物)反应器中，以反应物 A 的水溶液为原料，考察每小时生产一定量 F_B 克分子的化学产品 B 的问题．化学反应为



$$2A\rightarrow B, \tag{1.14}$$

且符合由实验建立的，以反应液的单位体积为基准的经验速率方程：

$$-\frac{dC_A}{dt}=8.4(C_A)^2=8.4[C_A^0(1-x_A)]^2\ \left(\frac{\text{克分子}}{\text{升-小时}}\right), \tag{1.15}$$

其中 C_A——反应器中 A 的浓度(克分子/升)，

C_A^0——原料中 A 的浓度(克分子/升)，

t——时间(小时)，

x_A——转换因子，反应物转化为产品的份数．

假设 A 的浓度在某一连续范围内时，原料溶液是可用的，并

且单位价格 p_A 由下列关系给出:

$$p_A=4(C_A^0)^{1.4}\ (\text{美元/升}), \tag{1.16}$$

连续搅拌槽式反应器(CSTR)的操作费用为

$$p_{\mathrm{CSTR}}=0.75(V)^{0.6}\ (\text{美元/小时}). \tag{1.17}$$

其中 V(升)是反应器的容积. 假定产品 B 的售价为 10 美元/克分子, 我们的问题是去决定: 原料溶液的速率 F_A^0(升/小时)、它的浓度 C_A^0、反应器的容积 V 和转换因子 x_A, 以达到最优操作. 也就是说, 使下列每小时的总收益为最大:

$$p_T=10F_B-p_AF_A^0-p_{\mathrm{CSTR}}\ (\text{美元/小时}). \tag{1.18}$$

反应器中物质的平衡给出

$$F_A^0C_A^0=F_A^0C_A^0(1-x_A)-\left(\frac{dC_A}{dt}\right)V. \tag{1.19}$$

由(1.14)得到

$$\frac{1}{2}F_A^0C_A^0x_A=F_B, \tag{1.20}$$

并由(1.15)和(1.19)得到

$$8.4C_A^0(1-x_A)^2V-F_A^0x_A=0, \tag{1.21}$$

于是, 设计问题成为

$$\max p_T=5F_A^0C_A^0x_A-4(C_A^0)^{1.4}F_A^0-0.75(V)^{0.6}, \tag{1.22}$$

受限制于约束(1.21). 这是变量 C_A^0, F_A^0, V, x_A 的非线性规划, 目标函数和约束函数都是非线性的. 注意, 这里是使目标函数极大化.

在整个这本书中, 我们将主要关心目标函数极小化的非线性规划. 这并不算任何限制, 因为每一个 $\max f(x)$型的问题, 通过考察 $\min \bar{f}(x)$, 其中 $\bar{f}(x)=-f(x)$, 总可以等价地进行分析和求解.

最后, 对以后各章所用的记号和术语说几句. 没有用特别的符号来表示向量. 向量的维数如不特别说明, 总可以从它出现的

式子中看出．所有的向量假定是列向量．一个向量的分量以下标表示，这样，x_1, x_2, $\cdots$, x_n 是 n 维向量 x 的分量．向量的上标用来区分不同的向量，这样，x^1, x^2, $\cdots$, x^m 表示 m 个不同的向量．为了避免混淆起见，出现实数的指数时将带有括弧，就是说 $(a)^2$ 是数 a 的平方．记号 x^T 用来表示一个行向量，即列向量的转置．实直线，即所有实数的集合用 R 来记，n 维实 Euclid 空间记为 R^n．向量 $x\in R^n$ 也常认为是 R^n 中的点 x．R^n 中特定的点常以它的坐标来表示，例如 $x^0=(-1, 2)$．记号 $x\geqslant 0$ 表示 x 的每个分量都是非负的．这样，如果 $0\in R^n$，那末 0 也是一个 n 维向量，它的每一个分量为零．记号 $x\neq 0$ 表示 x 至少有一个分量异于零．

仅仅用到具有实元素的矩阵．和向量类似，对矩阵也没有特别的符号(虽然一般用大写字母)．矩阵的维数从它出现的式子中唯一地确定．如果 A 是 $m\times n$ 阵(m 行 n 列)，那末 A 的转置 A^T 有 n 行 m 列．矩阵 A 的逆阵记为 A^{-1}，即 $AA^{-1}=A^{-1}A=I$，其中 I 是恒等(单位)阵．

向量 $x\in R^n$ 的模定义为

$$\|x\|=[(x_1)^2+(x_2)^2+\cdots+(x_n)^2]^{1/2}. \tag{1.23}$$

我们常用邻域的概念．集合

$$N_\delta(x^0)=\{x:x\in R^n,\ \|x-x^0\|<\delta\} \tag{1.24}$$

称为点 x^0 的一个(球形)邻域，这里 δ 是一正数．

矩阵的模在本书少数地方也用到，而它们是由向量诱导出的模．形式地说，如果 A 是一个 $m\times n$ 阵，x 是 n 维向量，那么

$$\|A\|=\sup\left\{\frac{\|Ax\|}{\|x\|}:\|x\|\neq 0\right\}=\sup\{\|Ax\|:\|x\|=1\}. \tag{1.25}$$

函数总是单值的；如以后将看到的．它们有时可取值 $+\infty$ 或 $-\infty$．在几处需要较深数学概念的地方，或是在书中定义它；或是在这个定义要求更多的背景材料时，我们就放弃完整性，要求读者去参看适当的文献．读者如对初等线性代数，实分析或拓扑不熟悉，希望去查阅这些方面的入门性教科书，诸如 Apostol[1]，Bartle[2]，Hall、Spencer[3] 与 Noble[4]．

参 考 文 献

1. Apostol, T. M., *Mathematical Analysis*, 2nd ed., Addison-Wesley Publishing Co., Reading, Mass., 1974.
2. Bartle, R. G., *The Elements of Real Analysis*, 2nd ed., John Wiley & Sons, New York, 1976.
3. Hall, D. W., and G. L. Spencer, *Elementary Topology*, John Wiley & Sons, New York, 1955.
4. Noble, B., *Applied Linear Algebra*, Prentice-Hall, Englewood Cliffs, N.J., 1969.

第 I 部分　分　析

如不作特别说明，求解一个非线性规划问题的涵义是求一个最优解向量 x^*，而不是求可能存在的所有最优解. 判定一个最优解 x^*，并研究它的性质，构成本书第 I 部分的中心论题，它处理非线性规划问题的某些分析方面的内容. 我们将会看到，如果向量 x 是最优解的候选者，它必需满足某些最优性必要条件. 但是，不幸的是可能有其他异于最优解的向量也满足这些条件. 这样，必要条件主要是在否定意义下使用的：如果向量 x 不满足条件，它就不是最优解. 因此，为了验证最优性，我们希望找出最优性的充分条件，当它与必要条件一起被满足时，就清楚地指明了所考虑的具体解向量的性质. 这两种类型的最优性条件是要稍详细讨论的第一个主题. 特别地，把经典的 Lagrange 乘子法扩充到具有不等式约束的最优化问题. 在一般的非线性规划中，向量可能满足一类最优性条件而不满足另一类. 最优性必要条件也是充分条件的规划是很重要的，因为这时在所考察的规划中，表示最优性必要条件的一组方程或不等式的解，必然是这规划所求的最优解.

称为凸规划的一类非线性规划就具有上述这样好的性质，它们以某种结构方式含有凸函数和凹函数. 这种凸规划特别便于分析. 与每一个这样的规划相联系，存在所谓对偶规划，与线性规划类似，它具有某些有趣的理论性质. 对偶性关系将借助于基于共轭函数的近代方式进行推导.

凸规划属于每个局部极小都是整体极小的那一类非线性规划. 由于有许多非凸的实际问题，我们还得考虑一个重要问题，就

是找出具有这种局部-整体最优性质的、较一般的非凸函数和非凸规划. 作为本书分析部分的结尾，选择了一些非线性规划，对它们阐明某些理论性的结果.

第 2 章

经典最优化——无约束和等式约束问题

求实函数的极值，即极小值或极大值的问题，在数学最优化中处于中心地位．我们从最简单的无约束问题开始这个极值课题，然后进入带有等式约束的极小和极大的论题．这里讨论经典的 Lagrange 乘子理论以及可微函数极值的某些必要和充分条件．这些课题的处理可以回溯到几个世纪以前，所以有“经典的”称呼．后面几章将讨论有不等式约束的最优化问题．这个问题得到的所有值得注意的结果可以算是“近代的”，因为它们是近二三十年来对不等式约束问题的强烈兴趣所引出的结果．所有“经典的”结果可以看成更一般的“近代”理论的特殊情况．我们先介绍经典的结果，是因为它们可以作为桥梁，把多数在第一和第二学年开设的大学微积分或实分析课程的内容，与数学规划更深一些的主题连接起来．此外，经典理论比近代理论在这样的意义下更简单，即诸如关于极值的充要条件等结果，不会象不等式约束情形那样被更复杂的要求弄得模糊不清．

2.1 无约束极值

考虑 R^n 中区域 D 上的一个实值函数 f，称 f 在一点 $x^*\in D$ 有**局部极小值**，如果存在一个实数 $\delta>0$，使得

$$f(x)\geqslant f(x^*) \tag{2.1}$$

对适合 $\|x-x^*\|<\delta$ 的一切 $x\in D$ 成立．同样可定义**局部极大值**，只要倒转(2.1)中的不等号．如果不等式(2.1)换为严格不等式

$$f(x)>f(x^*),\quad x\in D,\quad x\neq x^*, \tag{2.2}$$

就称 f 在 x^* 有**严格**局部极小值；而如果(2.2)的不等号反向，则得到严格局部极大值．如果(2.1)（或(2.2)）对一切 $x\in D$ 成立，称

函数 f 在 $x^*\in D$ 有整体极小值(严格整体极小值)；整体极大值(严格整体极大值)可类似定义．一个极值是指极大值或极小值．不是每个实函数都有极值，例如一个非零线性函数在 R^n 中没有极值．从定义明显得出：f 在 D 中的每一个整体极小(极大)值也是局部极小(极大)值．其逆一般是错的，读者容易用例子说明．但是在以后几章中我们将讨论一些函数，例如凸函数，它们具有值得注意的性质：每个局部极小值也是整体极小值．

设 $x\in D\subset R^n$ 是一点，实函数 f 在这点可微．我们知道若一个实值函数 f 在内点 $x\in D$ 可微，则在 x 处存在一阶偏导数．此外，若偏导数在 x 连续，则说 f 在 x 连续可微．同样，若 f 在 $x\in D$ 二次可微，则在那里存在二阶偏导数．若它们在 x 连续，则称 f 在 x 二次连续可微．定义 f 在点 x 的梯度为如下向量 $\nabla f(x)$：

$$\nabla f(x)=\left(\frac{\partial f(x)}{\partial x_1},\ \cdots,\ \frac{\partial f(x)}{\partial x_n}\right)^T. \tag{2.3}$$

类似地，若 f 在 x 二次可微，定义 f 在 x 的 Hesse 阵为如下的 $n\times n$ 对称阵 $\nabla^2 f(x)$：

$$\nabla^2 f(x)=\left[\frac{\partial^2 f(x)}{\partial x_i\partial x_j}\right],\quad i,\ j=1,\ \cdots,\ n. \tag{2.4}$$

本节我们讨论无约束函数极值的必要和充分条件，我们从叙述以下著名的结果开始．

定理 2.1(**必要条件**)

设 x^* 是 D 的内点，f 在这一点有局部极小值或局部极大值，若 f 在 x^* 可微，则

$$\nabla f(x^*)=0. \tag{2.5}$$

这个定理将作为定理 2.3 的一部分，重新叙述并证明．

现在转向局部极值的充分条件．

定理 2.2(**充分条件**)

设 x^* 是 D 的内点，在这点 f 二次连续可微，若

$$\nabla f(x^*)=0, \tag{2.6}$$

以及

$$z^T\nabla^2 f(x^*)z>0 \tag{2.7}$$

对一切非零向量 z 成立，则 f 在 x^* 有局部极小值．若(2.7)的不等号反向，则 f 在 x^* 有局部极大值．并且这些极值是严格局部极值．

这个定理可以用 f 的 Taylor 展开式来证明，留给读者去做．

在这两个定理中都用到函数在极值点 x^* 的性态．然而，如果我们研究函数在所讨论极值点的邻域中的性态，就可以给出关于局部极值的其他条件．

定理 2.3

设 x^* 是 D 的内点，且 f 在 D 上二次连续可微．x^* 为 f 的局部极小的必要条件是：

$$\nabla f(x^*)=0, \tag{2.8}$$

并对一切 z 成立

$$z^T\nabla^2 f(x^*)z\geqslant 0. \tag{2.9}$$

局部极小的充分条件是(2.8)成立，且对某个邻域 $N_\delta(x^*)$ 中的每个 x 和每个 $z\in R^n$ 有

$$z^T\nabla^2 f(x)z\geqslant 0. \tag{2.10}$$

如果(2.9)和(2.10)中不等号反向，定理可用于局部极大．

【证明】 设 f 在 x^* 有局部极小值，则对某个邻域 $N_\delta(x^*)\subset D$ 中的一切 x 成立

$$f(x)\geqslant f(x^*). \tag{2.11}$$

我们能把每个 $x\in N_\delta(x^*)$ 写为 $x=x^*+\theta y$，其中 θ 为一实数，y 是向量，其模 $\|y\|=1$．因此对充分小的 $|\theta|$ 成立

$$f(x^*+\theta y)\geqslant f(x^*). \tag{2.12}$$

对这样的 y，以 $F(\theta)=f(x^*+\theta y)$ 定义 F．于是(2.12)成为

$$F(\theta)\geqslant F(0), \tag{2.13}$$

它对满足 $|\theta|<\delta$ 的一切 θ 成立．

根据中值定理[1]，有

$$F(\theta)=F(0)+\nabla F(\lambda\theta)\theta, \tag{2.14}$$

其中 λ 是 0 和 1 之间的一个数．

若 $\nabla F(0)>0$，则由连续性假设，存在一个 $\varepsilon>0$，使得

$$\nabla F(\lambda\theta)>0 \tag{2.15}$$

对0和1之间的一切 λ 和适合 $|\theta|<\varepsilon$ 的一切 θ 成立. 因此可找到一个 $\theta<0$ 适合 $|\theta|<\delta$, 且

$$F(0)>F(\theta), \tag{2.16}$$

便得矛盾. 假若 $\nabla F(0)<0$, 将导致同样的矛盾. 因此

$$\nabla F(0)=y^T\nabla f(x^*)=0. \tag{2.17}$$

但是 y 是任意的非零向量, 因此有

$$\nabla f(x^*)=0. \tag{2.18}$$

现在转到二阶条件, 由 Taylor 定理,

$$F(\theta)=F(0)+\nabla F(0)\theta+\frac{1}{2}\nabla^2F(\lambda\theta)(\theta)^2,\quad 0<\lambda<1. \tag{2.19}$$

如果 $\nabla^2F(0)<0$, 那么由连续性, 存在 $\varepsilon'>0$ 使得

$$\nabla^2F(\lambda\theta)<0 \tag{2.20}$$

对0和1之间的一切 λ 以及适合 $|\theta|<\varepsilon'$ 的一切 θ 成立.

由于 $\nabla F(0)=0$, (2.20)对这样的 θ 为

$$F(\theta)<F(0), \tag{2.21}$$

从而发生矛盾. 因而

$$\nabla^2F(0)=y^T\nabla^2f(x^*)y\geqslant0. \tag{2.22}$$

由于这不等式对于只是模受到限制的一切 y 成立, 它必然对于一切向量 z 也成立. 这样便完成了定理第一部分的证明.

为证第二部分, 假设(2.8)和(2.10)成立, 但 x^* 不是局部极小值点. 则有一向量 $w\in N_\delta(x^*)$ 使 $f(x^*)>f(w)$. 设 $w=x^*+\theta y$, 而 $\|y\|=1$ 且 $\theta>0$. 根据 Taylor 定理,

$$f(w)=f(x^*)+\theta y^T\nabla f(x^*)+\frac{1}{2}(\theta)^2y^T\nabla^2f(x^*+\lambda\theta y)y, \tag{2.23}$$

其中 $0<\lambda<1$. 上面的假设导致

$$y^T\nabla^2f(x^*+\lambda\theta y)y<0, \tag{2.24}$$

由于 $x^*+\lambda\theta y\in N_\delta(x^*)$, 上式便与(2.10)相矛盾. 对局部极大的证明是类似的. ▌

定理 2.2 基于函数在点 x^* 的性态，提供了 f 在 x^* 有严格局部极值的充分条件. 我们将说明，容易找到不满足这些充分条件的极值的例子. 在定理 2.3 中我们有基于 f 在 x^* 的邻域中的性态的局部(不一定为严格)极值的充分条件. 最后，我们介绍一个也是基于 x^* 的邻域的严格局部极值的充分条件.

定理 2.4

设 x^* 是 D 的内点，且 f 为二次连续可微，若

$$\nabla f(x^*)=0, \tag{2.25}$$

且对 x^* 的邻域中任何 $x\neq x^*$ 和任何非零的 z 成立

$$z^T\nabla^2 f(x)z>0, \tag{2.26}$$

则 f 在 x^* 有严格局部极小值. 倒转(2.26)中的不等号，就得到严格局部极大值的充分条件.

这个定理的证明类似于前一定理，留给读者作为练习.

以一个简单的例子说明前面的定理.

例 2.1.1

设 $f(x)=(x)^{2p}$，其中 p 为正整数，D 为整个实直线. 梯度 ∇f 由 $\nabla f(x)=2p(x)^{2p-1}$ 给出. 在 $x=0$，梯度为零，就是说，原点满足定理 2.1 所述的极小或极大的必要条件.

Hesse 阵 $\nabla^2 f$ 为

$$\nabla^2 f(x)=(2p-1)(2p)(x)^{2p-2}. \tag{2.27}$$

对 $p=1$，$\nabla^2 f(0)=2$，即严格局部极小的充分条件(定理 2.2)满足.



图 2.1 函数 x^{2p} 在原点邻域中的图形

但是，如果取 $p>1$，那么 $\nabla^2 f(0)=0$，定理 2.2 的充分条件不满足，然而可以从图上看出 f 在原点有极小值. 另一方面，取原点的任意邻域，读者容易验证定理 2.3 局部极小(必要的和充分的)条件全满足，并且，由定理 2.4 可以断定极小值点 $x^*=0$ 是严格极小. 事实上，原点确是 f 的严格整体极小值点.

前面这些定理含有二阶条件，牵涉到由

$$z^T H(c) z = \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ij}(c) z_i z_j \tag{2.28}$$

给出的称为二次型的函数的性质，其中 $H=[h_{ij}]$ 是实对称阵．为了研究这种函数的符号，或者等价地，研究矩阵 $H(c)$ 的确定性，我们计算行列式 $d_k(c)$：

$$d_k(c) = \det \begin{bmatrix} h_{11}(c) & \cdots & h_{1k}(c) \\ \vdots & & \vdots \\ h_{k1}(c) & \cdots & k_{kk}(c) \end{bmatrix}, \quad k=1, \cdots, n. \tag{2.29}$$

若对 $k=1, \cdots, n$，成立 $d_k(c)>0$，二次型 $z^T H(c) z$ 就对一切非零的 z 为正，从而 $H(c)$ 是正定的．若对 $k=1, \cdots, n$，$d_k(c)$ 有 $(-1)^k$ 的符号，即 $d_k(c)$ 的值交替地为负和正，由(2.28)给出的二次型对一切非零的 z 为负，从而 $H(c)$ 是负定的．如果我们对 f 在某点 c 的性态感兴趣，就象定理 2.2 的充分条件只牵涉到矩阵 $\nabla^2 f(c)$ 那样，这些验算是有用的．但是，如果象定理 2.3 那样，要在 c 的一个邻域中确定二次型的符号，这些验算就行不通了．

2.2 等式约束极值和 Lagrange 方法

在上节我们讨论了没有其他附加条件时，函数在其定义域内部存在极值的必要和充分条件．这一节我们讨论的极值问题，其目标是求实值函数在含于函数定义域的一个特定范围内的极小值或极大值，而这容许范围是由有限个称为约束的方程来描述的．

考察求以 $D \subset R^n$ 为定义域的实值函数 f 的极小(或极大)值问题，受到的约束为

$$g_i(x)=0, \quad i=1, \cdots, m \ (m<n), \tag{2.30}$$

其中每个 g_i 是在 D 上定义的实值函数．约束方程个数小于变量个数这一假定将简化下面的讨论．所以问题是在由(2.30)中方程确定的范围内求 f 的极值．求解这种问题的第一个和最直观的方法包含这样的步骤，即利用(2.30)中的方程消去 m 个变量．消去的条件将在后面的隐函数定理中叙述，而证明可在大多数高等微

积分教本中找到(如参看[1, 8]). 这个定理假定函数 g_i 可微和 $n\times m$ Jacobi 阵 $[\partial g_i/\partial x_j]$ 的秩为 m. 约束方程中 m 个变量关于其余 $n-m$ 个变量的实际求解, 虽然不是不可能的, 却常常表明是困难的工作. 为了这个缘故, 并且显然这个方法把等式约束问题归约为一个等价的在前节讨论过的无约束问题, 我们就不进一步继续下去了.

另一种方法由 Lagrange 提出, 也是基于把一个约束问题变为一个无约束问题的想法. 数学规划的许多近代结果, 实际上是 Lagrange 方法的一种直接扩充和推广, 主要是针对不等式约束的.

我们先叙述著名的隐函数定理, 然后介绍一个结果, 沿着这结果所提供的方向, 可以把等式约束问题化为等价的无约束问题.

定理 2.5(隐函数定理)

设 ϕ_i 是定义在 D 上的实值函数, 在一开集 $D^1\subset D\subset R^{m+p}$ 上连续可微, 其中 $p>0$ 并且对 $i=1, \cdots, m$ 及 $(x^0, y^0)\in D^1$ 成立 $\phi_i(x^0, y^0)=0$. 假定 Jacobi 阵 $[\partial\phi_i(x^0, y^0)/\partial x_j]$ 的秩为 m. 那么存在一个邻域 $N_\delta(x^0, y^0)\subset D^1$、一个含 y^0 的开集 $D^2\subset R^p$ 和在 D^2 上连续可微的实值函数 ψ_k, $k=1, \cdots, m$, 使得下列条件满足:

$$x_k^0=\psi_k(y^0), \quad k=1, \cdots, m; \tag{2.31}$$

对每个 $y\in D^2$ 有

$$\phi_i(\psi(y), y)=0, \quad i=1, \cdots, m, \tag{2.32}$$

其中 $\psi(y)=(\psi_1(y), \cdots, \psi_m(y))$; 且对一切 $(x, y)\in N_\delta(x^0, y^0)$, Jacobi 阵 $[\partial\phi_i(x, y)/\partial x_j]$ 的秩为 m. 此外, 对 $y\in D^2$, $\psi_k(y)$ 的偏导数是下列线性方程组的解:

$$\sum_{k=1}^{m}\frac{\partial\phi_i(\psi(y), y)}{\partial x_k}\frac{\partial\psi_k(y)}{\partial y_j}=-\frac{\partial\phi_i(\psi(y), y)}{\partial y_j}, \quad i=1, \cdots, m. \tag{2.33}$$

在引进 Lagrange 方法之前, 先提出下述结果.

定理 2.6

设 f 和 $g_i(i=1, \cdots, m)$ 是 $D\subset R^n$ 上的实值函数, 且在一邻域

$N_\varepsilon(x^*)\subset D$ 上连续可微. 假设对 $N_\varepsilon(x^*)$中满足

$$g_i(x)=0,\quad i=1,\ \cdots,\ m \tag{2.34}$$

的一切点 x 来说, x^* 是 f 的一个局部极小或极大值点. 又假设 $g_i(x^*)$的 Jacobi 阵的秩为 m. 在这些假设下, f 在 x^* 的梯度是 g_i 在 x^* 的梯度的线性组合,就是说存在实数 λ_i^* 使得

$$\nabla f(x^*)=\sum_{i=1}^{m}\lambda_i^*\nabla g_i(x^*). \tag{2.35}$$

【证明】 经过对行的适当重新排列和编号, 总可以假定由 Jacobi 阵$[\partial g_i(x^*)/\partial x_j]$的前$m$行构成的 $m\times m$ 阵是非异的. 线性方程组

$$\sum_{i=1}^{m}\frac{\partial g_i(x^*)}{\partial x_j}\lambda_i=\frac{\partial f(x^*)}{\partial x_j},\quad j=1,\ \cdots,\ m \tag{2.36}$$

对 λ_i 有唯一解,记为 λ_i^*. 令 $\hat{x}=(x_{m+1},\ \cdots,\ x_n)$. 对于(2.34)在 x^* 应用定理 2.5, 于是存在实函数 $h_j(\hat{x})$和含 x^* 的开集 $\hat{D}\subset R^{n-m}$, 使得

$$x_j^*=h_j(\hat{x}^*),\quad j=1,\ \cdots,\ m \tag{2.37}$$

以及

$$f(x^*)=f(h_1(\hat{x}^*),\ \cdots,\ h_m(\hat{x}^*),\ x_{m+1}^*,\ \cdots,\ x_n^*). \tag{2.38}$$

作为上面式子的一个结果,由定理 2.1 可知, f 关于 $x_{m+1},\cdots,$ x_n 的偏导数在 x^* 必为零,这样,

$$\frac{\partial f(x^*)}{\partial x_j}=\sum_{k=1}^{m}\frac{\partial f(x^*)}{\partial x_k}\frac{\partial h_k(\hat{x}^*)}{\partial x_j}+\frac{\partial f(x^*)}{\partial x_j}=0,\quad j=m+1,\ \cdots,\ n. \tag{2.39}$$

从(2.33)对每个 $j=m+1,\ \cdots,\ n$ 有

$$\sum_{k=1}^{m}\frac{\partial g_i(x^*)}{\partial x_k}\frac{\partial h_k(\hat{x}^*)}{\partial x_j}=-\left(\frac{\partial g_i(x^*)}{\partial x_j}\right),\quad i=1,\ \cdots,\ m. \tag{2.40}$$

以 λ_i^* 乘(2.40)中每一个方程并相加, 得到

$$\sum_{i=1}^{m}\sum_{k=1}^{m}\lambda_i^*\frac{\partial g_i(x^*)}{\partial x_k}\frac{\partial h_k(\hat{x}^*)}{\partial x_j}+\sum_{i=1}^{m}\lambda_i^*\frac{\partial g_i(x^*)}{\partial x_j}=0,\ \ j=m+1,\ \cdots,\ n. \tag{2.41}$$

从(2.39)减去(2.41)并加整理, 就有

$$\sum_{k=1}^{m}\left[\frac{\partial f(x^*)}{\partial x_k}-\sum_{i=1}^{m}\lambda_i^*\frac{\partial g_i(x^*)}{\partial x_k}\right]\frac{\partial h_k(\hat{x}^*)}{\partial x_j}+\frac{\partial f(x^*)}{\partial x_j}-\sum_{i=1}^{m}\lambda_i^*\frac{\partial g_i(x^*)}{\partial x_j}$$
$$=0,\quad j=m+1,\ \cdots,\ n. \tag{2.42}$$

但由(2.36)，方括弧中的式子为零，所以

$$\frac{\partial f(x^*)}{\partial x_j}-\sum_{i=1}^{m}\lambda_i^*\frac{\partial g_i(x^*)}{\partial x_j}=0,\quad j=m+1,\ \cdots,\ n. \tag{2.43}$$

最后的表达式与(2.36)一起得到要证的结果． ▌

在局部极值点，求极小或极大的函数的梯度与约束函数的梯度之间，存在上面定理所表示的关系，它导致 Lagrange 式 $L(x,\ \lambda)$ 的下述表达式：

$$L(x,\ \lambda)=f(x)-\sum_{i=1}^{m}\lambda_i g_i(x), \tag{2.44}$$

其中 λ_i 称为 **Lagrange 乘子**.

Lagrange 方法包含有将一个等式约束极值问题化为求 Lagrange 式的逗留点问题．这可由以下定理看出．

定理 2.7

设 $f,\ g_i,\ i=1,\ \cdots,\ m$ 满足定理 2.6 的假设，则存在一个乘子向量 $\lambda^*=(\lambda_1^*,\ \cdots,\ \lambda_m^*)^T$，使

$$\nabla L(x^*,\ \lambda^*)=0. \tag{2.45}$$

【证明】 从定理 2.6 和(2.44)给出的 L 的定义直接推出． ▌

在文献中有上面两个定理的一些不同证明，例如[2, 6, 7]，我们选择基于隐函数定理的证明，这是因为它不需要其他背景材料．前已指出，本节的定理可以看成某些近代更一般结果的特殊情况．下一章我们也将介绍它们，并将给出沿不同路线的证明．

定理 2.7 提供了有等式约束的 f 的极值的必要条件．和前一节一样，现在转到讨论这种极值的充分条件．记号 $\nabla_\xi^j\phi(\xi,\ \eta)$ 用来表示 ϕ 关于 ξ 的 j 次导数(当 $j=1$ 时就省略上标)，于是有下述定理．

定理 2.8

设 $f,\ g_1,\ \cdots,\ g_m$ 是 R^n 上二次连续可微的实值函数．若存在向量 $x^*\in R^n$ 和 $\lambda^*\in R^m$，使得

$$\nabla L(x^*, \lambda^*)=0, \tag{2.46}$$

并且，对每个非零向量 $z\in R^n$，只要满足

$$z^T\nabla g_i(x^*)=0, \quad i=1, \cdots, m, \tag{2.47}$$

便有

$$z^T\nabla_x^2 L(x^*, \lambda^*)z>0, \tag{2.48}$$

则在限制 $g_i(x)=0$, $i=1, \cdots, m$ 下，f 在 x^* 有严格局部极小值。若(2.48)的不等式反向，则 f 在 x^* 有严格局部极大值。

【证明】[6] 假定 x^* 不是严格局部极小值点，则必存在一个邻域 $N_\delta(x^*)$ 和一个收敛于 x^* 的序列 $\{z^k\}$，使 $z^k\in N_\delta(x^*)$, $z^k\neq x^*$，且对每个 $z^k\in\{z^k\}$ 有

$$g_i(z^k)=0, \quad i=1, \cdots, m, \tag{2.49}$$

$$f(x^*)\geqslant f(z^k). \tag{2.50}$$

令 $z^k=x^*+\theta^k y^k$，其中 $\theta^k>0$ 及 $\|y^k\|=1$，序列 $\{\theta^k, y^k\}$ 有子序列收敛于 $(0, \bar{y})$，而 $\|\bar{y}\|=1$. 根据中值定理[1]，对这子序列的每个 k，我们有

$$g_i(z^k)-g_i(x^*)=\theta^k(y^k)^T\nabla g_i(x^*+\eta_i^k\theta^k y^k)=0, \quad i=1, \cdots, m, \tag{2.51}$$

其中 η_i^k 是 0 和 1 之间的一个数，且

$$f(z^k)-f(x^*)=\theta^k(y^k)^T\nabla f(x^*+\xi^k\theta^k y^k)\leqslant 0, \tag{2.52}$$

其中 ξ^k 也是 0 和 1 之间的一个数.

以 θ^k 除(2.51)和(2.52)，并令 $k\to\infty$，得到

$$(\bar{y})^T\nabla g_i(x^*)=0, \quad i=1, \cdots, m \tag{2.53}$$

$$(\bar{y})^T\nabla f(x^*)\leqslant 0. \tag{2.54}$$

从 Taylor 定理有

$$L(z^k, \lambda^*)=L(x^*, \lambda^*)+\theta^k(y^k)^T\nabla_x L(x^*, \lambda^*)$$
$$+\frac{1}{2}(\theta^k)^2(y^k)^T\nabla_x^2 L(x^*+\eta^k\theta^k y^k, \lambda^*)y^k, \tag{2.55}$$

其中 $1>\eta^k>0$.

根据(2.44), (2.46), (2.49)及(2.50)并用 $\frac{1}{2}(\theta^k)^2$ 除(2.55)，得到

$$(y^k)^T \nabla_x^2 L(x^* + \eta^k \theta^k y^k, \lambda^*) y^k \leqslant 0, \tag{2.56}$$

令 $k \to \infty$，从上式得到

$$(\bar{y})^T \nabla_x^2 L(x^*, \lambda^*) \bar{y} \leqslant 0. \tag{2.57}$$

由于 $\bar{y} \neq 0$ 并满足(2.47)，这就完成了证明. ∎

上一定理中叙述的充分条件，涉及到在线性约束下决定一个二次型的符号. 这一工作可以由 Mann[5] 的一个结果来完成. 设 $A = [\alpha_{ij}]$ 是一个 $n \times n$ 实对称阵，且 $B = [\beta_{ij}]$ 为 $n \times m$ 实阵，以 M_{pq} 记由阵 M 仅仅保留前 p 行和前 q 列的元素所得到的矩阵.

定理 2.9

假设 $\det[B_{mm}] \neq 0$，那么二次型

$$\sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_{ij} \xi_i \xi_j \tag{2.58}$$

对一切适合

$$\sum_{i=1}^{n} \beta_{ij} \xi_i = 0, \quad j = 1, \cdots, m \tag{2.59}$$

的非零向量 ξ 为正，当且仅当

$$(-1)^m \det \begin{bmatrix} A_{pp} & B_{pm} \\ B_{pm}^T & 0 \end{bmatrix} > 0 \tag{2.60}$$

对 $p = m+1, \cdots, n$ 成立. 类似地，(2.58)对一切适合(2.59)的非零向量 ξ 为负，当且仅当

$$(-1)^p \det \begin{bmatrix} A_{pp} & B_{pm} \\ B_{pm}^T & 0 \end{bmatrix} > 0 \tag{2.61}$$

对 $p = m+1, \cdots, n$ 成立.

这个定理的证明可以在[3]或[5]中找到.

假设 $n \times m$ 的 Jacobi 阵 $[\partial g_i(x^*)/\partial x_j]$ 的秩为 m，且变量的标号使得

$$\det \begin{bmatrix} \dfrac{\partial g_1(x^*)}{\partial x_1} & \cdots & \dfrac{\partial g_m(x^*)}{\partial x_1} \\ \vdots & & \vdots \\ \dfrac{\partial g_1(x^*)}{\partial x_m} & \cdots & \dfrac{\partial g_m(x^*)}{\partial x_m} \end{bmatrix} \neq 0, \tag{2.62}$$

那么有下面的结果.

推论 2.10

设 f, g_1, $\cdots$, g_m 是二次连续可微的实值函数, 如果存在向量 $x^*\in R^n$, $\lambda^*\in R^m$ 使得

$$\nabla L(x^*, \lambda^*)=0, \tag{2.63}$$

并且如果

$$(-1)^m \det\begin{bmatrix} \frac{\partial^2 L(x^*, \lambda^*)}{\partial x_1\partial x_1} & \cdots & \frac{\partial^2 L(x^*, \lambda^*)}{\partial x_1\partial x_p} & \frac{\partial g_1(x^*)}{\partial x_1} & \cdots & \frac{\partial g_m(x^*)}{\partial x_1} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial^2 L(x^*, \lambda^*)}{\partial x_p\partial x_1} & \cdots & \frac{\partial^2 L(x^*, \lambda^*)}{\partial x_p\partial x_p} & \frac{\partial g_1(x^*)}{\partial x_p} & \cdots & \frac{\partial g_m(x^*)}{\partial x_p} \\ \frac{\partial g_1(x^*)}{\partial x_1} & \cdots & \frac{\partial g_1(x^*)}{\partial x_p} & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial g_m(x^*)}{\partial x_1} & \cdots & \frac{\partial g_m(x^*)}{\partial x_p} & 0 & \cdots & 0 \end{bmatrix} > 0 \tag{2.64}$$

对 $p=m+1$, $\cdots$, n 成立, 那么 f 在 x^* 有严格局部极小值, 使得

$$g_i(x^*)=0, \quad i=1, \cdots, m. \tag{2.65}$$

【证明】 从定理 2.8 和定理 2.9 直接得到. 】

对于严格局部极大值的类似结果, 在 (2.64) 中改 $(-1)^m$ 为 $(-1)^p$ 便得.

例 2.2.1

考察问题

$$\max f(x_1, x_2)=x_1x_2, \tag{2.66}$$

受限制于约束

$$g(x_1, x_2)=x_1+x_2-2=0. \tag{2.67}$$

首先作 Lagrange 式

$$L(x, \lambda)=x_1x_2-\lambda(x_1+x_2-2). \tag{2.68}$$

其次取 $\nabla L(x^*, \lambda^*)=0$:

$$\frac{\partial L(x^*, \lambda^*)}{\partial x_1} = x_2^* - \lambda^* = 0, \tag{2.69}$$

$$\frac{\partial L(x^*, \lambda^*)}{\partial x_2} = x_1^* - \lambda^* = 0, \tag{2.70}$$

$$\frac{\partial L(x^*, \lambda^*)}{\partial \lambda} = -x_1^* - x_2^* + 2 = 0. \tag{2.71}$$

上面三个方程的解为

$$x_1^* = x_2^* = \lambda^* = 1. \tag{2.72}$$

因此点 $(x^*, \lambda^*) = (1, 1, 1)$ 满足定理 2.7 中所述的极大值点的必要条件.

根据定理 2.6, 在最优点 ∇f 和 ∇g 必定线性相关, 这明显地从图 2.2 看出, 在那点 $\nabla f(x^*)$ 实际上等于 $\nabla g(x^*)$.

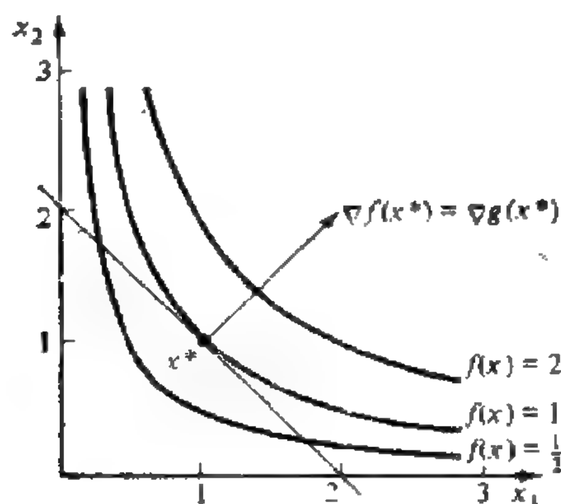


图 2.2 有约束的极大

转向充分条件, 计算 $\nabla_x^2 L(x^*, \lambda^*)$:

$$\frac{\partial^2 L(x^*, \lambda^*)}{\partial x_1 \partial x_1} = 0, \quad \frac{\partial^2 L(x^*, \lambda^*)}{\partial x_1 \partial x_2} = 1, \quad \frac{\partial^2 L(x^*, \lambda^*)}{\partial x_2 \partial x_2} = 0. \tag{2.73}$$

所以

$$z^T \nabla_x^2 L(x^*, \lambda^*) z = (z_1, z_2) \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = 2z_1 z_2. \tag{2.74}$$

根据定理 2.8, 必须对适合 $z^T \nabla g(x^*) = 0$ 的一切 $z \neq 0$, 决定函数 $2z_1z_2$ 的符号.

由于

$$\frac{\partial g(x^*)}{\partial x_1}=1, \quad \frac{\partial g(x^*)}{\partial x_2}=1, \tag{2.75}$$

上面的条件等价于 $z_1+z_2=0$. 代入(2.74), 得到

$$z^T\nabla_x^2 L(x^*, \lambda^*)z=-2(z_1)^2<0, \tag{2.76}$$

因此(1, 1)是严格局部极大值点.

最后, 可以检验推论 2.10 给出的充分条件. 这时 $p=2$, 且

$$(-1)^2\det\begin{bmatrix}0 & 1 & 1\\ 1 & 0 & 1\\ 1 & 1 & 0\end{bmatrix}=2>0, \tag{2.77}$$

因而验证了前面的推论. ▌

已对更一般的约束极值问题导出了类似于定理 2.3 的二阶必要条件[4, 6]. 在等式约束的情形, 这些条件几乎是无约束问题的二阶必要和充分条件的直接推广.

在定理 2.6 中我们假定了 Jacobi 阵 $[\partial g_i(x^*)/\partial x_j]$ 的秩为 $m(<n)$, 它等于约束方程的个数. 在结束这一章时, 我们叙述定理 2.6 的一点推广, 它对于 Jacobi 阵的秩数不要求条件.

定理 2.11

设 f 和 g_i, $i=1, \cdots, m$, 是区域 $D\subset R^n$ 上连续可微的实值函数. 若对 x^* 的一个邻域内满足

$$g_i(x)=0, \quad i=1, \cdots, m \tag{2.78}$$

的一切点 x 来说, x^* 是 f 的局部极小或极大值点, 则存在 $m+1$ 个不全为零的实数 λ_0^*, λ_1^*, $\cdots$, λ_m^* 使得

$$\lambda_0^*\nabla f(x^*)-\sum_{i=1}^{m}\lambda_i^*\nabla g_i(x^*)=0. \tag{2.79}$$

下一章要介绍既有等式又有不等式约束的极值的更一般的结果, 这个定理可看作它的一个推论. 从这些结果我们可以得出以下结论: 若 x^* 是一个局部极值点, 则 $n\times(m+1)$ 增广 Jacobi 阵 $[\partial f(x^*)/\partial x_j, \partial g_i(x^*)/\partial x_j]$ 的秩小于 $(m+1)$. 此外, 可以证明, 若以上矩阵的秩等于 Jacobi 阵 $[\partial g_i(x^*)/\partial x_j]$ 的秩, 则 $\lambda_0^*\neq 0$, 并可规范化为 $\lambda_0^*=1$. 但是, 若增广 Jacobi 阵的秩大于 Jacobi 阵

$[\partial g_i(x^*)/\partial x_j]$ 的秩，则 λ_0^* 必定为零. 例如能行集只含有一个点时，可以发生这个情况。

例 2.2.2

$$\min f(x)=x \tag{2.80}$$

受限制于

$$g_1(x)=(x)^2=0, \tag{2.81}$$

能行集只包含一点 $x=0$. 在这点上，Jacobi 阵 $[dg_1/dx]$ 的秩为零，而增广 Jacobi 阵 $[df/dx, dg_1/dx]$ 的秩为 1. 从(2.79)有

$$\lambda_0^*-\lambda_1^*\cdot 0=0, \tag{2.82}$$

即 $\lambda_0^*=0$.

练　习

2.A. 证明定理2.2.

2.B. 对下面的函数，求出满足极值必要条件的点：

$$f(x)=\frac{x_1+x_2}{3+(x_1)^2+(x_2)^2+x_1x_2}. \tag{2.83}$$

试通过检验充分条件确定这些点的性质.

2.C. 研究 R^3 的原点是否为函数

$$f(x)=\alpha(x_1)^2e^{x_2}+(x_2)^2e^{x_3}+(x_3)^2e^{x_1} \tag{2.84}$$

的极值点，其中 α 为参数.

2.D. 对于下面的函数，在满足极值必要条件的一切点中，求极小值点：

$$f(x)=[(x_2)^2-x_1]^2. \tag{2.85}$$

2.E. 证明定理2.4.

2.F. 给定了线性方程组

$$Ax=b, \tag{2.86}$$

其中 A 是 $m\times n$ 实矩阵，b 是 m 维向量，$x\in R^n$ 是未知量. 假定 $m<n$ 且 A 的秩是 m. 使 $\frac{1}{2}x^Tx$ 为极小的(2.86)的解记为 x^*，求出用 A 和 b 表示的 x^* 的显式解答. 说明 $x^*=(41, -9, -1)^T$ 确实是

$$2x_1-x_2+5x_3=86 \tag{2.87}$$

$$x_1-2x_3=43 \tag{2.88}$$

的这样一个解，以检验你所得的结果.

2.G. 如果 $g_i(x^*)$ 的梯度是线性相关的，必要条件(2.35)不一定成立，举一个数值例子来说明这事实.

2.H. 分别用下列方法求点 $\bar{x}=(1, 0)$ 到曲线

$$4x_1-(x_2)^2=0 \tag{2.89}$$

的最短距离：(a)直接消去一个变量，(b)Lagrange 方法．这两种情况是否得到相同的解．

参 考 文 献

1. BARTLE, R. G., *The Elements of Real Analysis*, 2nd ed., John Wiley & Sons, New York, 1976.
2. BELTRAMI, E. J., "A Constructive Proof of the Kuhn-Tucker Multiplier Rule," *J. Math. Anal. & Appl.*, 26, 297–306 (1967).
3. DEBREU, G., "Definite and Semidefinite Quadratic Forms," *Econometrica*, 20, 295–300 (1952).
4. FIACCO, A. V., "Second Order Sufficient Conditions for Weak and Strict Constrained Minima," *SIAM J. Appl. Math.*, 16, 105–108 (1968).
5. MANN, H. B., "Quadratic Forms with Linear Constraints," *Amer. Math. Monthly*, 50, 430–433 (1943).
6. MCCORMICK, G. P., "Second Order Conditions for Constrained Minima," *SIAM J. Appl. Math.*, 15, 641–652 (1967).
7. PHIPPS, C. G., "Maxima and Minima Under Restraint," *Amer. Math. Monthly*, 59, 230–235 (1952).
8. RUDIN, W., *Principles of Mathematical Analysis*, 2nd ed., McGraw-Hill Book Co., New York, 1964.

第 3 章

约束极值的最优性条件

上一章处理的问题局限于无约束或等式约束问题. 本章我们开始讨论含有不等式和等式约束的数学规划问题. 注意, 把不等式引入最优化问题, 标志着最优化"经典"时代的结束和数学规划"现代"理论的开始. 不等式约束很少是严格的不等式, 而可以作为等式或严格不等式被满足. 不等式的这个特点使最优性条件的分析处理复杂化,但足以补偿这一点的是,利用不等式约束能够表达极为丰富的一类问题.

最优性条件问题的某些处理可以在文献中找到, 每种处理的特点是由施加于所含函数的假定来刻划的. 在这些处理的大部分中,所叙述的最优性条件以这种或那种方式与 Lagrange 式的概念有关. 如果目标函数和约束函数是可微的(或二次连续可微), 那么 Lagrange 式既能较方便地处理, 又不损失许多一般性, 本章通篇使用这一假定. 放松可微性假定,往往导致另一种最优性条件, 它们最好表示为另一些问题的最优性条件,例如求 Lagrange 式的鞍点,或求解所谓对偶规划.

3.1 不等式约束极值的一阶必要条件

我们来叙述本章所讨论的最一般的数学规划问题, 着手导出不等式与等式约束极值问题的一阶必要条件, 其中仅包含一阶导数. 问题:

(P)
$$\min f(x) \tag{3.1}$$

受限制于约束

$$g_i(x)\geqslant 0, \quad i=1, \cdots, m, \tag{3.2}$$

$$h_j(x)=0, \quad j=1, \cdots, p. \tag{3.3}$$

假定函数 f, g_1, $\cdots$, g_m, h_1, $\cdots$, h_p 在某开集 $D\subset R^n$ 中定义并可

微. 以 $X\subset D$ 表示问题(P)的能行集，即满足(3.2)与(3.3)的所有 $x\in D$ 的集合. 能行集中的点称为能行点. 如前，以 $N_\delta(x^0)$ 表示点 x^0 的半径 δ 的球形邻域.

点 $x^*\in X$ 称为问题(P)的局部极小值点或(P)的局部解，如果存在正数 $\hat{\delta}$ 使

$$f(x)\geqslant f(x^*) \tag{3.4}$$

对所有 $x\in X\cap N_{\hat{\delta}}(x^*)$ 成立. 如果(3.4)对所有 $x\in X$ 成立，则 x^* 称为问题(P)的整体极小值点(整体解).

每个在 x^* 邻域中的点 x 可表示为 x^*+z，这里，当且仅当 $x\neq x^*$ 时 z 为非零向量. 向量 $z\neq 0$ 称为 x^* 的能行方向向量，如果存在 $\delta_1>0$，使 $(x^*+\theta z)\in X\cap N_{\delta_1}(x^*)$ 对于所有的 $0\leqslant\theta<\delta_1/\|z\|$ 成立(见图 3.1).

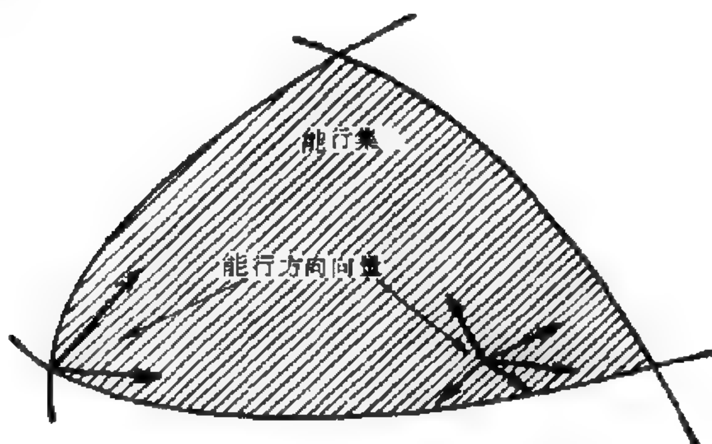


图 3.1 能行方向向量

能行方向向量在许多数值最优化算法中是重要的. 当前我们对它感兴趣，简单的理由是，如果 x^* 是问题(P)的局部解，z 是能行方向向量，那么对于充分小的正数 θ，必有 $f(x^*+\theta z)\geqslant f(x^*)$. 让我们用约束函数 g_i 与 h_j 刻划能行方向向量的特征.

定义

$$I(x^*)=\{i: g_i(x^*)=0\}. \tag{3.5}$$

假设对于某个 $k\in I(x^*)$ 与 x^* 点的一个能行方向向量 z，成立 $z^T\nabla g_k(x^*)<0$. 按照可微性假设，我们有

$$g_k(x^*+\theta z)=g_k(x^*)+\theta z^T\nabla g_k(x^*)+\theta\varepsilon_k(\theta), \tag{3.6}$$

其中 $\varepsilon_k(\theta)$ 当 $\theta\to 0$ 时趋于零. 如果 θ 足够小, 则 $z^T\nabla g_k(x^*)+\varepsilon_k(\theta)<0$, 并且因为 $g_k(x^*)=0$, 我们得到 $g_k(x^*+\theta z)<0$ 对于所有充分小的 $\theta\geqslant 0$ 成立, 从而与 z 是 x^* 的能行方向向量这个事实相矛盾. 因此, 对于所有 $i\in I(x^*)$, 必须有 $z^T\nabla g_i(x^*)\geqslant 0$. 同理可证, 对于能行方向向量, 也必须对 $j=1, \cdots, p$ 有 $z^T\nabla h_j(x^*)=0$. 现在定义

$$Z^1(x^*)=\{z: z^T\nabla g_i(x^*)\geqslant 0,\ i\in I(x^*);\quad z^T\nabla h_j(x^*)=0,\ j=1, \cdots, p\}. \tag{3.7}$$

通过前面的讨论可见, 如果 z 是 x^* 的能行方向向量, 则 $z\in Z^1(x^*)$. 一个集 $K\subset R^n$ 称为锥, 如果对于每一个非负数 α, $x\in K$ 蕴涵 $\alpha x\in K$. 集 $Z^1(x^*)$ 显然是一个锥. 它也称为 X 在 x^* 的**线性化锥**[1,17], 因为它通过在 x^* 线性化约束函数而生成. 让我们定义另一"线性化"集合 $Z^2(x^*)$ 以备以后需要:

$$Z^2(x^*)=\{z: z^T\nabla f(x^*)<0\}. \tag{3.8}$$

如果 $z\in Z^2(x^*)$, 可以证明存在足够接近于 x^* 的点 $x=x^*+\theta z$, 使得 $f(x^*)>f(x)$.

由 Minkowski、Farkas[13] 得到并以后者命名的下述引理, 是接下去所需要的. 这个引理的证明在下一章给出.

引理 3.1(Farkas 引理)

设 A 是给定的 $m\times n$ 实矩阵, b 是给定的 n 维向量. 对所有满足 $Ay\geqslant 0$ 的向量 y 成立不等式 $b^Ty\geqslant 0$, 其充要条件是存在 m 维向量 $\rho\geqslant 0$ 使 $A^T\rho=b$.

图 3.2 说明了对于 3×2 阵 A 的 Farkas 引理. 向量 A_1, A_2, A_3 是矩阵 A 的行向量, 考虑由与 A 的每个行向量成锐角的所有向量 y 组成的集 Y. Farkas 引理指出, b 与每个 $y\in Y$ 成锐角的充要条件是 b 能表示成 A 的行向量的非负线性组合. 在图 3.2 中, b^1 是满足这些条件的向量, 而 b^2 却不是.

如在前面一章那样, 我们现在定义关于问题(P)的 Lagrange 式

$$L(x, \lambda, \mu)=f(x)-\sum_{i=1}^{m}\lambda_i g_i(x)-\sum_{j=1}^{p}\mu_j h_j(x), \tag{3.9}$$

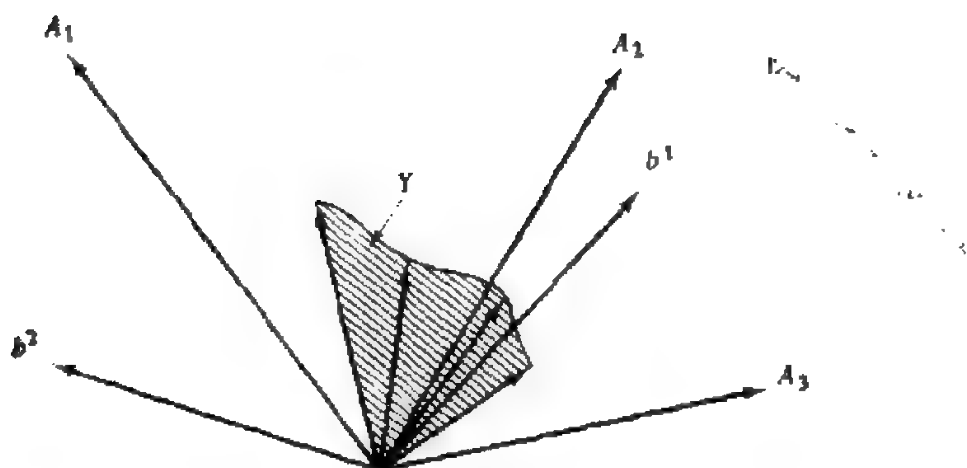


图 3.2 对于 3×2 矩阵 A 的 Farkas 引理的解释

并可证明下列定理.

定理 3.2

设 $x^0\in X$，则 $Z^1(x^0)\cap Z^2(x^0)=\emptyset$ 的充要条件为：存在向量 λ^0 和 μ^0，使得

$$\nabla_x L(x^0,\ \lambda^0,\ \mu^0)=\nabla f(x^0)-\sum_{i=1}^{m}\lambda_i^0\nabla g_i(x^0)-\sum_{j=1}^{p}\mu_j^0\nabla h_j(x^0)=0, \tag{3.10}$$

$$\lambda_i^0 g_i(x^0)=0,\quad i=1,\ \cdots,\ m, \tag{3.11}$$

$$\lambda^0\geqslant 0. \tag{3.12}$$

【证明】 集 $Z^1(x^0)$是非空的，因为原点总是属于它；$Z^1(x^0)\cap Z^2(x^0)$为空集的充要条件是：对满足

$$z^T\nabla g_i(x^0)\geqslant 0,\quad i\in I(x^0), \tag{3.13}$$

$$z^T\nabla h_j(x^0)=0,\quad j=1,\ \cdots,\ p \tag{3.14}$$

的每一个 z，我们有

$$z^T\nabla f(x^0)\geqslant 0. \tag{3.15}$$

我们可将(3.14)写成两个不等式

$$z^T\nabla h_j(x^0)\geqslant 0,\quad j=1,\ \cdots,\ p, \tag{3.16}$$

$$z^T[-\nabla h_j(x^0)]\geqslant 0,\quad j=1,\ \cdots,\ p. \tag{3.17}$$

由引理 3.1 可知，(3.15)对满足(3.13)、(3.16)、(3.17)的所有向量 z 成立的充要条件是：存在向量 $\lambda^0\geqslant 0$、$\mu^1\geqslant 0$ 和 $\mu^2\geqslant 0$，使得

$$\nabla f(x^0)=\sum_{i\in I(x^0)}\lambda_i^0\nabla g_i(x^0)+\sum_{j=1}^{p}(\mu_j^1-\mu_j^2)\nabla h_j(x^0). \qquad (3.18)$$

对于 $i\in I(x^0)$，令 $\lambda_i^0=0$，$\mu^0=\mu^1-\mu^2$，我们断言 $Z^1(x^0)\cap Z^2(x^0)$ 为空集的充要条件是(3.10)至(3.12)成立. ▌

我们希望扩充定理 2.6 中给出的关于等式约束问题之解的必要条件，对此，条件(3.10)～(3.12)当然是自然的候选者. 它们确实成为一般数学规划问题(P)的最优性的必要条件，如果我们能保证在(P)的局部解 x^* 处集 $Z^1(x^*)\cap Z^2(x^*)$ 为空集.

有兴趣的读者可以尝试在这点配上具有等式与不等式约束的简单的数学规划，得到它们的解，并对这些解点，构造集 $Z^1(x)$ 与 $Z^2(x)$. 对于大部分问题，他将发现在问题的解点处 $Z^1(x)\cap Z^2(x)$ 确实是空的，因此 Lagrange 式条件(3.10)至(3.12)在那些点成立. 然而并不总是这种情形，这从下述取自[25]的例子可见.

例 3.1.1

考虑 R^2 中的约束(见图 3.3)

$$g_1(x)=(1-x_1)^3-x_2\geqslant 0, \qquad (3.19)$$

$$g_2(x)=x_1\geqslant 0, \qquad (3.20)$$

$$g_3(x)=x_2\geqslant 0. \qquad (3.21)$$

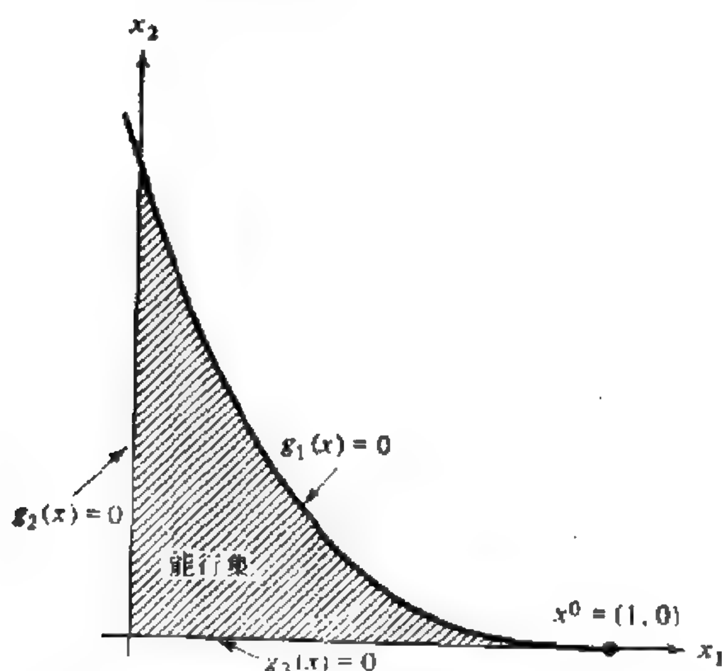


图 3.3 不具有 Lagrange 乘子的约束与最优点的例子

点 $x^0=(1, 0)$是能行的，容易验证

$$Z^1(x^0)=\{(z_1, z_2):z_2=0\}, \tag{3.22}$$

令 $f(x)=-x_1$，可以看到 x^0 是问题

$$\min -x_1 \tag{3.23}$$

受限制于约束(3.19)～(3.21)的解．在 x^0，

$$Z^2(x^0)=\{(z_1, z_2):z_1>0\}, \tag{3.24}$$

且 $Z^1(x^0)\cap Z^2(x^0)$非空，因此不存在满足(3.10)～(3.12)的 λ^0． ▮

在关于问题(P)的 Lagrange 函数的定义中引进目标函数的一个乘子，可能导出关于最优性的弱的必要条件，而不要求在解处集 $Z^1(x)\cap Z^2(x)$成为空集．令弱 **Lagrange 式** $\tilde{L}$ 定义为

$$\tilde{L}(x, \lambda, \mu)=\lambda_0 f(x)-\sum_{i=1}^{m}\lambda_i g_i(x)-\sum_{j=1}^{p}\mu_j h_j(x). \tag{3.25}$$

Fritz John 在他的 1948 年的开创性工作[22]中，通过弱 Lagrange 式，仅假定所包含的函数有连续的一次偏导数，就叙述并证明了不等式约束数学规划(没有等式约束)的必要条件．Jonh 的条件后来为 Mangasarian 与 Fromovitz 扩充到有等式与不等式约束的问题，即如本节开头所述的问题(P)．我们即将叙述这些条件，对仅有不等式约束的稍简单的情况给出证明．对不等式及等式情况的完整证明可以在[27]中找到．首先我们需要下面称为"择一定理"的结果．

定理 3.3

设 $\tilde{A}$ 是 $m\times n$ 实矩阵，则或者存在 n 维向量 x，使

$$\tilde{A}x<0, \tag{3.26}$$

或者存在m维非零向量 u，使

$$u^T\tilde{A}=0, \quad u\geqslant 0, \tag{3.27}$$

但二者不能同时成立．

【证明】 假设存在 x 与 u 使(3.26)和(3.27)同时满足，则我们同时有 $u^T\tilde{A}x<0$ 及 $u^T\tilde{A}x=0$，得出矛盾．现在如果不存在 x 满足(3.26)，这意味着我们找不到一个负数 w，使对每个 $x\in R^n$ 满足

$$\widetilde{A}_i x = \sum_{j=1}^{n} \tilde{a}_{ij} x_j \leqslant w, \quad i=1, \cdots, m. \tag{3.28}$$

令 $y=(w, x)^T$, $b=(1, 0, \cdots, 0)^T \in R^{n+1}$, $A=[e, -\widetilde{A}]$, 其中 $e=(1, \cdots, 1)^T \in R^m$, 借助前述 Farkas 引理, 我们断言存在$m$维向量 $u \geqslant 0$, 使

$$\sum_{i=1}^{m} u_i = 1, \tag{3.29}$$

$$\sum_{i=1}^{m} u_i a_{ij} = 0, \quad j=1, \cdots, n, \tag{3.30}$$

因此 u 是(3.27)的解. ∎

关于另一些择一定理, 读者可参考 Mangasarian[26]. 现在要叙述在上面定义的弱 Lagrange 式基础上的必要条件. 然而在证明中, 我们假设问题中不出现等式约束 $h_j(x)=0$. 于是下面的定理便回到 Fritz John 最初的那一个定理[22].

定理 3.4

假设 f, g_1, $\cdots$, g_m, h_1, $\cdots$, h_p 在包含X的开集上连续可微, 若 x^* 是问题(P)的解, 则存在向量 $\lambda^*=(\lambda_0^*, \lambda_1^*, \cdots, \lambda_m^*)^T$ 与 $\mu^*=(\mu_1^*, \cdots, \mu_p^*)^T$, 使得

$$\nabla_x \widetilde{L}(x^*, \lambda^*, \mu^*) = \lambda_0^* \nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^{p} \mu_j^* \nabla h_j(x^*) = 0, \tag{3.31}$$

$$\lambda_i^* g_i(x^*) = 0, \quad i=1, \cdots, m, \tag{3.32}$$

$$(\lambda^*, \mu^*) \neq 0, \quad \lambda^* \geqslant 0. \tag{3.33}$$

【证明】 对下列问题:

$$\min f(x) \tag{3.34}$$

受限制于

$$g_i(x) \geqslant 0, \quad i=1, \cdots, m, \tag{3.35}$$

我们考虑它的解 x^* 的必要条件. 这条件就是存在向量 λ^*, 使得

$$\lambda_0^* \nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0, \tag{3.36}$$

$$\lambda_i^* g_i(x^*) = 0, \quad i=1, \cdots, m, \tag{3.37}$$

$$\lambda^* \neq 0, \quad \lambda^* \geqslant 0. \tag{3.38}$$

若对所有 i 成立 $g_i(x^*)>0$，则 $I(x^*)=\emptyset$. 选取

$$\lambda_0^*=1, \quad \lambda_1^*=\lambda_2^*=\cdots=\lambda_m^*=0,$$

(3.36)至(3.38)与 $\nabla f(x^*)=0$ 一起成立.

现在设 $I(x^*)\neq\emptyset$，则对满足

$$z^T\nabla g_i(x^*)>0, \quad i\in I(x^*) \tag{3.39}$$

的每个 z，我们不能有

$$z^T\nabla f(x^*)<0. \tag{3.40}$$

这结果是从上面看到的事实推出的，即若存在 z 满足(3.39)，则能找到充分小的 δ，使对于任何 $0<\theta<\delta$，$x=x^*+\theta z$ 满足

$$g_i(x)>0, \quad i=1, \cdots, m; \tag{3.41}$$

即 x 是能行的. 若(3.40)也成立，则

$$f(x)<f(x^*), \tag{3.42}$$

与 x^* 是极小值点矛盾. 这样，不等式组(3.39)与(3.40)无解，由定理 3.3，存在一非零向量 $\lambda^*\geqslant0$，使得

$$\lambda_0^*\nabla f(x^*)+\sum_{i\in I(x^*)}\lambda_i^*[-\nabla g_i(x^*)]=0. \tag{3.43}$$

对 $i\notin I(x^*)$，令 $\lambda_i^*=0$，经过整理我们从(3.43)得到

$$\lambda_0^*\nabla f(x^*)-\sum_{i=1}^{m}\lambda_i^*\nabla g_i(x^*)=0, \tag{3.44}$$

并且显然

$$\lambda_i^*g_i(x^*)=0, \quad i=1, \cdots, m. \tag{3.45}$$

∎

如果 λ_0^* 是正的，定理 3.4 的条件(3.31)至(3.33)变成定理 3.2 的条件(3.10)至(3.12). 反之，定理 3.2 的条件自然蕴涵定理 3.4 当 $\lambda_0^*=1$ 时的那些条件.

例 3.1.2

再次考虑例 3.1.1 所讨论的极小化问题. 令 $\lambda_0^*=0$, $\lambda_1^*=1$, $\lambda_2^*=0$，以及 $\lambda_3^*=1$，我们注意到在 $x^*=(1, 0)$，Fritz John 的必要条件是满足的. ∎

这个例子也说明了定理 3.4 的主要弱点：事实上，代入 $\lambda_0^*=0$，条件(3.31)至(3.33)在(1, 0)对任何可微的目标函数 f 都满足，

不管它在这点是否有局部极值.

读者可能感到奇怪，既然我们基本的数学规划问题容易转换成只包含等式或不等式约束的等价问题，为什么它要同时具有等式与不等式约束. 假定在问题中有不等式 $g(x)\geqslant 0$. 令 y 是附加变量，我们可以写出等价的等式

$$g(x)-(y)^2=0. \tag{3.46}$$

相反，如果我们有等式约束 $h(x)=0$，它能被两个不等式所代替:

$$h(x)\geqslant 0, \tag{3.47}$$

$$h(x)\leqslant 0. \tag{3.48}$$

这样，以增加变量个数或约束个数为代价，等式-不等式类型问题能转换成只含单一类型约束的等价问题. 正如 Mangasarian、Fromovitz[37]关于 Fritz John 条件所指出的那样，这种转换在某些场合是有益的，但也可能使某些结果大大减弱.

假设我们将问题(P)的每个等式约束写成

$$h_j(x)=g_{m+j}(x)\geqslant 0, \quad j=1,\ \cdots,\ p, \tag{3.49}$$

$$-h_j(x)=g_{m+p+j}(x)\geqslant 0, \quad j=1,\ \cdots,\ p. \tag{3.50}$$

从而将问题(P)转换成(3.34)、(3.35)的形式. 选取

$$\lambda_0^*=\cdots=\lambda_m^*=0, \qquad \lambda_{m+1}^*=\cdots=\lambda_{m+p}^*=\lambda_{m+p+1}^*=\cdots=\lambda_{m+2p}^*=1,$$

我们发现，事实上条件(3.36)至(3.38)对于每个能行点满足，而不必是最优点.

在 1951 年，Kuhn 与 Tucker 在他们的基本著作[35]中给出了不等式约束的数学规划的必要条件，它比 John 的条件更强，用一个称为约束品性的正则性条件限制约束函数，它们能保证乘子 λ_0 确是正的. 这样，John 的必要条件的扩充形式便成为等价于条件(3.10)至(3.12)的形式. 对约束函数施加限制，以保证在数学规划的解处 Lagrange 乘子的存在，这种限制的类型已经成为努力深入研究的主题. 除了 Kuhn-Tucker 原来的约束品性之外已提出了另一些正则性条件. 我们在这里提到的著作有 Abadie[1]，Arrow、Hurwicz、Uzawa[2]，Beltrami[6]，Canon、Cullum、Polak[10]，Evans[13]，Gould、Tolle[17]，Guignard[19]，Karlin[23]，Mangasarian、

Fromovitz[27], Slater[35], Varaiya[36]. 在下面的讨论中我们一般遵循 Abadie, Gould 与 Tolle 以及 Varaiya 的著作. 对于各种约束品性的广泛讨论, 可参见 Bazaraa、Goode、Shetty[5] 和 Gould、Tolle[15]给出的概述.

实际上我们应把这些正则性条件称为"一阶"约束品性, 以区别于我们将在下一节提到的与二阶必要条件相联系的其他一些约束品性. 然而在不发生混淆时, 我们就简单地把它们称为约束品性.

我们通过引进非空集 $A\subset R^n$ 在点 $x\in A$ 的切锥的概念, 来开始讨论约束品性. 用 $\tilde{S}(A, x)$表示包含集 $\{a-x:a\in A\}$ 的所有闭锥的交, 则集 A 在点 x 的闭切锥 $S(A, x)$定义为

$$S(A, x)=\bigcap_{k=1}^{\infty}\tilde{S}(A\cap N_{1/k}(x), x), \tag{3.51}$$

其中 $N_{1/k}(x)$是 x 的半径为 $1/k$ 的球形邻域, k 是自然数. 下述引理刻划了 $S(A, x)$的特征.

引理 3.5

向量 z 包含在 $S(A, x)$的充要条件是: 存在一列收敛于 x 的向量$\{x^k\}\subset A$, 并存在一列非负数 $\{\alpha^k\}$, 使序列$\{\alpha^k(x^k-x)\}$收敛于 z.

【证明】[7] 设 $z\in S(A, x)$, 则对每个 $k=1, 2, \cdots$, 有 $z\in\tilde{S}(A\cap N_{1/k}(x), x)$. 按定义,

$$\tilde{S}(A\cap N_{1/k}(x), x)=\mathrm{cl}\{\alpha(y-x):\alpha\geqslant 0,\ y\in A\cap N_{1/k}(x)\},$$
$$k=1, 2, \cdots, \tag{3.52}$$

这里 cl 表示 R^n 中集合的闭包算子. 选取任一正数列$\{\varepsilon^k\}$使其收敛于 0, 并考虑向量 $z(\varepsilon^k)\in\{\alpha(y-x):\alpha\geqslant 0,\ y\in A\cap N_{1/k}(x)\}$ 使

$$\|z(\varepsilon^k)-z\|<\varepsilon^k, \tag{3.53}$$

则由(3.52), 它们可写作

$$z(\varepsilon^k)=\alpha(\varepsilon^k)(y(\varepsilon^k)-x),\quad \alpha(\varepsilon^k)\geqslant 0,\quad y(\varepsilon^k)\in A\cap N_{1/k}(x). \tag{3.54}$$

令 $k=1, 2, \cdots$, 我们产生一列包含于 A 且收敛于 x 的向量

$y(\varepsilon^1)$, $y(\varepsilon^2)$, …, 以及一非负数列 $\alpha(\varepsilon^1)$, $\alpha(\varepsilon^2)$, …, 使得由 (3.53) 及 (3.54), 序列 $\{\alpha(\varepsilon^k)(y(\varepsilon^k)-x)\}$ 收敛于 z. 反之, 假设存在一列收敛于 x 的向量 $\{x^k\}\subset A$ 和一列非负数 $\{\alpha^k\}$, 使 $\{\alpha^k(x^k-x)\}$ 收敛于 z, 令 p 是任一自然数, 则存在一自然数 K, 使当 $k\geqslant K$ 时就得到 $x^k\in A\cap N_{1/p}(x)$, 或

$$\alpha^k(x^k-x)\in\bar{S}(A\cap N_{1/p}(x),\ x),\quad k\geqslant K,\tag{3.55}$$

由于 $\bar{S}$ 是闭的, 有

$$z\in\bar{S}(A\cap N_{1/p}(x),\ x).\tag{3.56}$$

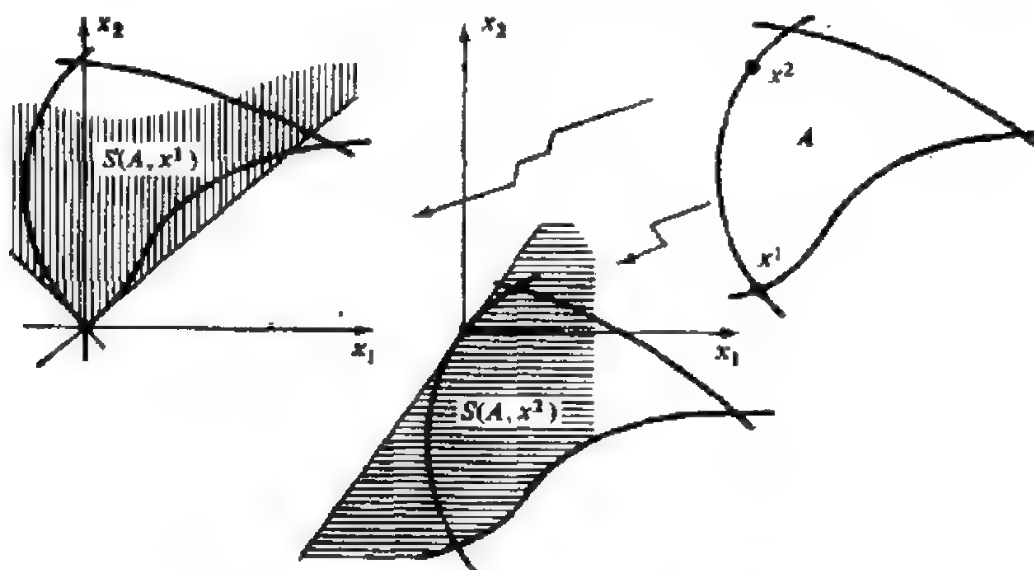


图 3.4　任意集 A 与它的切锥

因为最后表达式对任意自然数 p 成立, 推知

$$z\in\bigcap_{p=1}^{\infty}\bar{S}(A\cap N_{1/p}(x),\ x)=S(A,\ x).\tag{3.57}$$

]

借助上述引理, 可能给出非空集 A 在 x 点处切锥 $S(A,\ x)$ 的另一种描述. 首先注意到向量 $w=0$ 对每一 A 及 x 均在 $S(A,\ x)$ 内. 令 w 是一单位向量, 即 $\|w\|=1$. 设存在一个收敛于 x 的点列 $\{x^k\}\subset A$, $x^k\neq x$, 且

$$\lim_{k\to\infty}\frac{x^k-x}{\|x^k-x\|}=w.\tag{3.58}$$

我们可以说向量序列 $\{x^k\}$ 在 w 方向收敛于 x, 于是集 A 在 x 点的

切锥，包含上面得到的w的非负倍数的所有向量．因此由$w\in S(A, x)$推出存在一列$\{x^k\}\subset A$，在w方向收敛于x．察看切锥的另一种方式是这样的：通过把A的每一元素减去x以平移集A，令$\{x^k\}$是平移后集合中的序列，$x^k\neq 0$，收敛于原点．构造通过原点和x^k的射线序列，这些射线趋向于一条属于$S(A, x)$的射线．取所有这样的序列形成的所有射线，其全体将是A在x的切锥．这种构造如图3.5及下面例子所示．

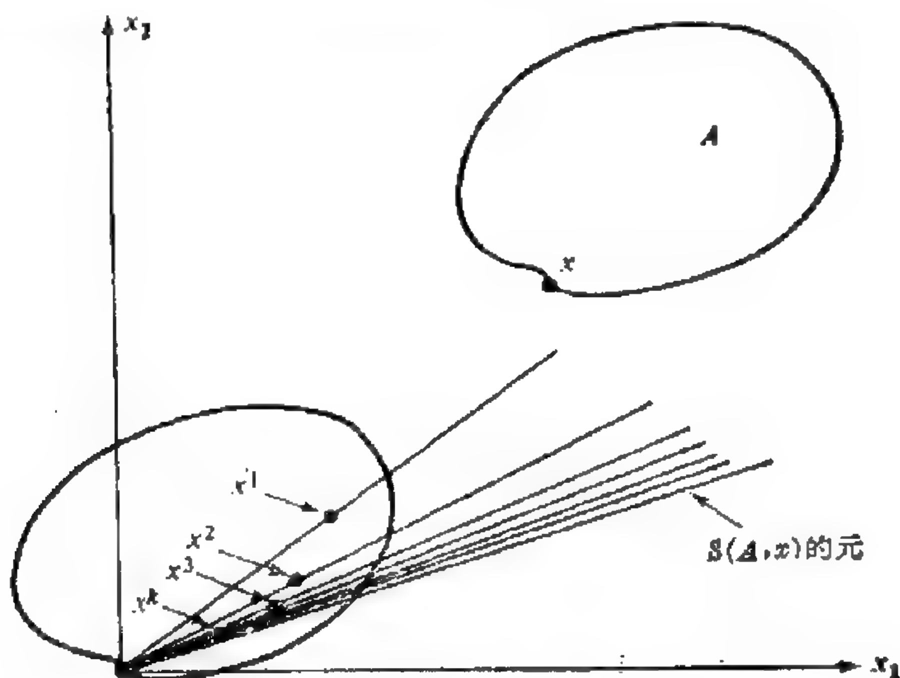


图 3.5　构造集A的切锥

例 3.1.3

设

$$B=\{(x_1, x_2):(x_1-4)^2+(x_2-2)^2\leqslant 1\}, \tag{3.59}$$

即B是中心在(4, 2)半径为1的闭球．让我们来找B在边界点，比方说$x=(4-\sqrt{3}/2, 3/2)$的切锥．首先通过把每一点减去x来平移B，得到球

$$B^1=\left\{(x_1, x_2):\left(x_1-\frac{\sqrt{3}}{2}\right)^2+\left(x_2-\frac{1}{2}\right)^2\leqslant 1\right\}. \tag{3.60}$$

在B^1的边界上取一列收敛于原点的序列$\{x^k\}$，我们产生一列射线，收敛于一直线，它实际上就是由B^1的边界定义的曲线在原点

处的普通切线，这直线满足

$$\frac{\sqrt{3}}{2}x_1+\frac{1}{2}x_2=0. \tag{3.61}$$

对于在 B^1 的内部而收敛于原点的所有序列重复上述过程，我们得到 B 在 x 的切锥为

$$S(B,\ x)=\left\{(x_1,\ x_2):\frac{\sqrt{3}}{2}x_1+\frac{1}{2}x_2\geqslant 0\right\}. \tag{3.62}$$

这是一个容易的练习，去说明在这情况下 $S(B,\ x)$ 与

$$g(x_1,\ x_2)=-(x_1-4)^2-(x_1-2)^2+1\geqslant 0 \tag{3.63}$$

在 $x=(4-\sqrt{3}/2,\ 3/2)$ 的线性化锥相符. ∎

我们接着将用到集 $A\subset R^n$ 的正法锥的概念，它由所有满足

$$x^Ty\geqslant 0,\ \text{对于所有}\ y\in A \tag{3.64}$$

的向量 $x\in R^n$ 组成，记作 A'. 我们用正法锥的名称以区别于“负”法锥或极锥. 它是用(3.64)中不等号反向来定义的. 下面用到的法锥的一个重要性质是：给定两个集合 $A_1\subset R^n$ 和 $A_2\subset R^n$，则

$$A_1\subset A_2\ \text{蕴涵}\ A_2'\subset A_1'. \tag{3.65}$$

切锥与正法锥在建立强最优性条件时起着核心的作用. 我们用下述引理开始把它们联系起来.

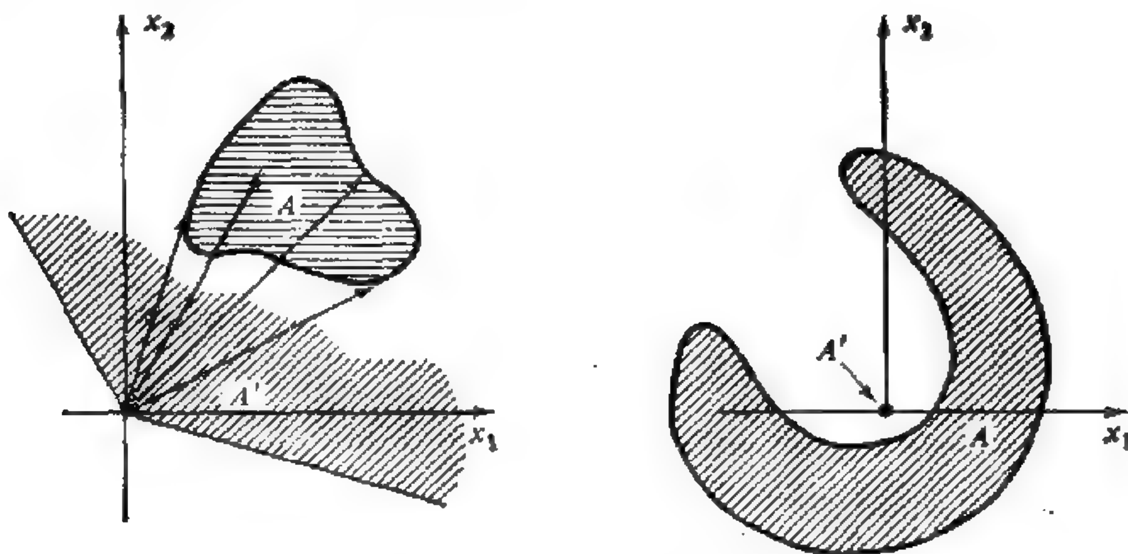


图3.6 正 法 锥

引理3.6

设 $x^0\in X$，则当且仅当

$$\nabla f(x^0)\in(Z^1(x^0))' \tag{3.66}$$

时，集 $Z^1(x^0)\cap Z^2(x^0)$ 为空集.

【证明】 集 $Z^1(x^0)\cap Z^2(x^0)$ 为空集，当且仅当对每个 $z\in Z^1(x^0)$ 有 $z^T\nabla f(x^0)\geqslant 0$，于是 $\nabla f(x^0)$ 包含在 $Z^1(x^0)$ 的正法锥内. ∎

引理3.7

设 x^0 是问题(P)的解，则

$$\nabla f(x^0)\in(S(X,\ x^0))'. \tag{3.67}$$

【证明】 我们必须证明，对每个 $z\in S(X,\ x^0)$，有 $z^T\nabla f(x^0)\geqslant 0$. 设 $z\in S(X,\ x^0)$，则由引理3.5，存在收敛于 x^0 的一序列 $\{x^k\}\in X$ 和一非负数列 $\{\alpha^k\}$，使得 $\{\alpha^k(x^k-x^0)\}$ 收敛于 z. 若 f 在 x^0 可微，我们可写出

$$f(x^k)=f(x^0)+(x^k-x^0)^T\nabla f(x^0)+\varepsilon\|x^k-x^0\|, \tag{3.68}$$

其中 ε 是当 $k\to\infty$ 时趋向于零的函数. 因此

$$\alpha^k[f(x^k)-f(x^0)]=(\alpha^k(x^k-x^0))^T\nabla f(x^0)+\varepsilon\|\alpha^k(x^k-x^0)\|. \tag{3.69}$$

因为 $x^k\in X$，x^0 是局部极小值点，令 $k\to\infty$，则得 $\varepsilon\|\alpha^k(x^k-x^0)\|\to 0$，且表示式 $\alpha^k[f(x^k)-f(x^0)]$ 收敛于非负极限. 这样

$$\lim_{k\to\infty}(\alpha^k(x^k-x^0))^T\nabla f(x^0)=z^T\nabla f(x^0)\geqslant 0, \tag{3.70}$$

即 $\nabla f(x^0)\in(S(X,\ x^0))'$. ∎

现在我们可以叙述并证明本节的主要结果，它是比定理3.4中给出的要更强的一组必要条件. 下面叙述的条件可看作 Kuhn-Tucker 最优性必要条件的直接扩充.

定理3.8(广义 Kuhn-Tucker 必要条件)

令 x^* 是问题(P)的解，并设

$$(Z^1(x^*))'=(S(X,\ x^*))', \tag{3.71}$$

则存在向量 $\lambda^*=(\lambda_1^*,\ \cdots,\ \lambda_m^*)^T$ 与 $\mu^*=(\mu_1^*,\ \cdots,\ \mu_p^*)^T$，使得

$$\nabla f(x^*)-\sum_{i=1}^{m}\lambda_i^*\nabla g_i(x^*)-\sum_{j=1}^{p}\mu_j^*\nabla h_j(x^*)=0, \tag{3.72}$$

$$\lambda_i^* g_i(x^*)=0, \quad i=1, \cdots, m, \tag{3.73}$$

$$\lambda^*\geqslant 0. \tag{3.74}$$

【证明】 设 x^* 是问题(P)的解，由引理3.7，$\nabla f(x^*)\in(S(X, x^*))'$. 若 $(Z^1(x^*))'=(S(X, x^*))'$，则 $\nabla f(x^*)\in(Z^1(x^*))'$. 由引理3.6，集 $Z^1(x^*)\cap Z^2(x^*)$ 是空的．再由定理3.2，条件(3.72)至(3.74)成立．】

注意，在问题(P)的解 x^* 处，下式

$$Z^1(x^*)=S(X, x^*) \tag{3.75}$$

蕴涵上述定理的假设．Gould、Tolle[17]，对一个比我们这里略为一般的问题曾用(3.71)作为约束品性．在他们的著作中，附加限制 $x\in D$ 被施于向量 x 上，其中 D 是 R^n 的任意子集．在他们的情形中，能行集 X 是 D 与满足(3.2)与(3.3)的 x 的集合的交集．他们证明了，(3.71)不仅是存在乘子 λ^*、μ^* 使(3.72)至(3.74)成立的充分条件，而且在一定意义上是必要的．三重组(g, h, D)称为在 x^* 是 **Lagrange 正则的**，当且仅当对每个在 x^* 有局部约束极小值的可微目标函数 f，存在向量 λ^*、μ^* 满足(3.72)至(3.74)．后来在[17]中证明了，(g, h, D)在 x^* 是 Lagrange 正则的，当且仅当条件(3.71)成立．实质上，我们在定理3.8中已经证明了，(3.71)确实是存在满足(3.72)至(3.74)的乘子 λ^*、μ^* 的充分条件．为了证明这条件也是必要的，Gould 与 Tolle 证明了对每个 $y\in(S(X, x^*))'$，存在具有局部约束极小值的可微函数 f，使 $\nabla f(x^*)=y$，由 Lagrange 正则性与引理3.6，我们得到 $y\in(Z^1(x^*))'$. 因此

$$(S(X, x^*))'\subset(Z^1(x^*))'. \tag{3.76}$$

要求读者证明，对每个能行点 $\hat{x}$ 有

$$(Z^1(\hat{x}))'\subset(S(X, \hat{x}))', \tag{3.77}$$

因此等式满足．

例 3.1.4

考虑下列非线性规划问题：

$$\min f(x) = x_1, \tag{3.78}$$

受限制于

$$g_1(x) = 16 - (x_1 - 4)^2 - (x_2)^2 \geqslant 0, \tag{3.79}$$

$$h_1(x) = (x_1 - 3)^2 + (x_2 - 2)^2 - 13 = 0. \tag{3.80}$$

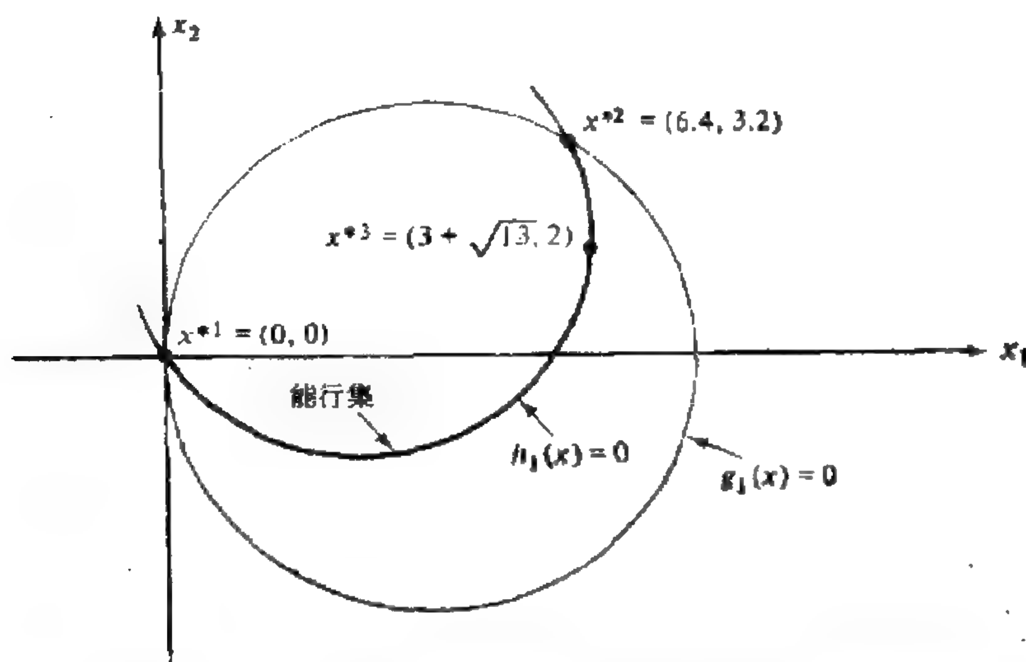


图 3.7　例 3.1.4 的能行集和满足 Kuhn-Tucker 条件的点

可以从图 3.7 中看到 $f(x)$ 在 $x^{*1} = (0,\ 0)$ 与 $x^{*2} = (6.4,\ 3.2)$ 有局部极小值，在这两点，$I(x^{*1}) = I(x^{*2}) = \{1\}$. 在第一点，

$$Z^1(x^{*1}) = \left\{(z_1,\ z_2) : z_1 \geqslant 0,\ z_2 = -\frac{3}{2} z_1\right\}, \tag{3.81}$$

并且这集合也是 $S(X,\ x^{*1})$，这可用简单的构造来验证. 现在

$$Z^2(x^{*1}) = \{(z_1,\ z_2) : z_1 < 0\}, \tag{3.82}$$

因此 $Z^1(x^{*1}) \cap Z^2(x^{*2}) = \emptyset$. 对 $\lambda_1^* = \frac{1}{8}$ 和 $\mu_1^* = 0$, Kuhn-Tucker 条件(3.72)至(3.74)满足. 在第二点，

$$Z^1(x^{*2}) = \left\{(z_1,\ z_2) : z_1 \geqslant 0,\ z_2 = -\frac{17}{6} z_1\right\}, \tag{3.83}$$

$$Z^2(x^{*2}) = \{(z_1,\ z_2) : z_1 < 0\}, \tag{3.84}$$

并且又有 $Z^1(x^{*2}) \cap Z^2(x^{*2}) = \emptyset$. 读者可验证，在第二个局部极小值点所需的乘子是

$$\lambda_1^* = \frac{3}{40} \quad 和 \quad \mu_1^* = \frac{1}{5}.$$

其实还有另外的成立 Kuhn-Tucker 必要条件的能行点. 令 $x^{*3} = (3+\sqrt{13}, 2)$, 读者能证明 $Z^1(x^{*3}) \cap Z^2(x^{*3}) = \emptyset$, 且相应的乘子是 $\lambda_1^* = 0$ 和 $\mu_1^* = \frac{\sqrt{13}}{26}$. 检查图 3.7, 发现 x^{*3} 不是我们问题的解, 而正好是下列问题:

$$\max f(x) = x_1 \tag{3.85}$$

在约束(3.79)至(3.80)下的解. ▮

到现在为止, 关于问题(P)中所包含的函数的类型, 除了可微性, 没有作特殊的假定. 关于这些函数的进一步假定将导致 Kuhn-Tucker 条件的特殊形式. 我们仅给出一种特殊情况. 在很多应用中, 出现在规划中的变量 x_j 要求是非负的. 假设在约束(3.2)和(3.3)之外我们还要求

$$x \geqslant 0, \tag{3.86}$$

对这种情形的必要条件可以叙述如下:

定理 3.9

令 x^* 是问题(P)在附加约束(3.86)下的解, 设(3.75)成立, 则存在向量

$$\lambda^* = (\lambda_1^*, \cdots, \lambda_m^*) \quad 和 \quad \mu^* = (\mu_1^*, \cdots, \mu_p^*),$$

使

$$\nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^{p} \mu_j^* \nabla h_j(x^*) \geqslant 0, \tag{3.87}$$

$$\lambda_i^* g_i(x^*) = 0, \quad i = 1, \cdots, m, \quad \lambda^* \geqslant 0, \tag{3.88}$$

$$(x^*)^T \left[\nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^{p} \mu_j \nabla h_j(x^*) \right] = 0. \tag{3.89}$$

这个定理的证明留给读者. 注意, 在这里能行集 X 由满足(3.2)、(3.3)和(3.86)的所有 x 组成. 对于求 f 的极大值而不是象(P)中那样求极小值的问题, 或者某些约束形为 $g_i(x) \leqslant 0$ 的问题, 容易借助于考虑在原来函数前添加负号, 以得到解的必要条

件，即在第一种情形求$-f(x)$的极小值，而在第二种情形受限制于约束$-g_i(x)\geqslant 0$.

对于一般非线性规划问题，必要且充分的约束品性(3.71)的验证是一件几乎不可能的事情．幸而实际上约束品性通常是成立的，所以假定乘子λ_1^*, …, λ_m^*, μ_1^*, …, μ_p^*的存在是很合理的，$\lambda_1^*, \cdots, \lambda_m^*$, $\mu_1^*, \cdots, \mu_p^*$称为**广义 Lagrange 乘子**或**Kuhn-Tucker 乘子**，它们满足定理3.8的一阶条件．我们叙述另外两个约束品性以结束这节，它们蕴涵(3.75)，在这意义下它们比已给出的那个要更强．

Kuhn-Tucker[26]原来的约束品性，在点$x^0\in X$要求任何向量$z\in Z^1(x^0)$与包含在X中的一可微弧相切．即对每个$z\in Z^1(x^0)$，存在一函数α，它的定义域是$[0, \varepsilon]\subset R$且值域在$R^n$中，使得$\alpha(0)=x^0$、对$0\leqslant\theta\leqslant\varepsilon$有$\alpha(\theta)\in X$、$\alpha$在$\theta=0$可微以及对某个正数$\lambda$有

$$\frac{d\alpha(0)}{d\theta}=\lambda z. \tag{3.90}$$

另一个约束品性是由 Mangasarian、Fromovitz[27]引进的．设g_i和h_j在点x^0分别是可微与连续可微的，如果向量$\nabla h_j(x^0)$, $j=1, \cdots, p$是线性独立的，且存在向量$z\in R^n$，使得

$$z^T\nabla g_i(x^0)<0, \quad i\in I(x^0), \tag{3.91}$$

$$z^T\nabla h_j(x^0)=0, \quad j=1, \cdots, p, \tag{3.92}$$

则约束品性成立．当约束函数是某种类型时适用的进一步的约束品性将在下一章中给出．

关于最优性的一阶必要条件在文献中曾广泛地讨论，除了本章和第一章列举的参考文献外，读者可以在下列作者的著作中找到最优性条件的分析：Bazaraa[3]，Bazaraa、Goode[4]，Braswell、Marban[8, 9]，Canon、Cullum、Polak[10]，Dubovitskii、Milyutin[11]，Gamkrelidze[16]，Halkin、Neustadt[20]，Neustadt[31]，Ritter[32, 33, 34]和 Wilde[38]．更特殊的最优性条件，主要是关于凸和广义凸非线性规划的，也将在后面几章中讨论．

3.2 二阶最优性条件

本节讨论关于问题(P)的含有二阶导数的最优性条件. 我们从二阶必要条件开始，它补充了前节的 Kuhn-Tucker 条件，然后叙述并证明在问题(P)中最优性的充分条件，扩充了第二章的相应结果.

读者可以回忆起，在我们关于一阶必要条件的推导中，它们是作为“集合的某个交集是空集”这一事实的结果得到的，对于二阶的情形，可以遵循同样的途径. 有一个这种形式的非常优美的推导，它属于 Dubovitskii、Milyutin[21]. 在 Messerli、Polak[29]的类似的推导中，得到了基于弱 Lagrange 式的某个一般的二阶必要条件，它可用于具有最低限度正则性条件的问题. 然而，为了使结论与证明简短些，我们将牺牲一般性. 下列结果是 McCormick[28] 得到的，他也导出了关于问题(P)的最优性充分条件，这将在后面给出.

在下面的讨论中，我们假定出现在问题(P)中的函数 f, $g_1, \cdots, g_m$, $h_1, \cdots, h_p$ 是二次连续可微的. 首先将叙述二阶约束品性. 设 $x\in X$，定义

$$\hat{Z}^1(x)=\{z: z^T\nabla g_i(x)=0,\ i\in I(x),\ z^T\nabla h_j(x)=0,\ j=1,\cdots,p\}. \tag{3.93}$$

如果每个非零的 $z\in \hat{Z}^1(x^0)$ 与包含于 X 的边界中的一条二次可微弧相切，则称为在 $x^0\in X$ 处二阶约束品性成立，即对每个 $z\in \hat{Z}^1(x^0)$，存在一个定义在$[0, \varepsilon]\subset R$ 上、值域在 R^n 中的二次可微函数 α，使得 $\alpha(0)=x^0$ 且对 $0\leqslant\theta\leqslant\varepsilon$ 有

$$g_i(\alpha(\theta))=0, \quad i\in I(x^0), \quad h_j(\alpha(\theta))=0, \quad j=1,\cdots,p \tag{3.94}$$

以及对某个正数 λ 有

$$\frac{d\alpha(0)}{d\theta}=\lambda z. \tag{3.95}$$

于是我们有以下定理.

定理 3.10(二阶必要条件)

设 x^* 是问题(P)的解，并设存在满足(3.72)至(3.74)的向量 $\lambda^*=(\lambda_1^*, \cdots, \lambda_m^*)^T$ 和 $\mu^*=(\mu_1^*, \cdots, \mu_p^*)^T$，进一步设在 x^* 处二阶约束品性成立，则对 $z\neq 0$ 且 $z\in \hat{Z}^1(x^*)$，我们有

$$z^T\left[\nabla^2 f(x^*)-\sum_{i=1}^m \lambda_i^*\nabla^2 g_i(x^*)-\sum_{j=1}^p \mu_j^*\nabla^2 h_j(x^*)\right]z\geqslant 0. \quad (3.96)$$

【证明】 令 $z\neq 0$ 且 $z\in \hat{Z}^1(x^*)$，并令 $\alpha(\theta)$ 是在二阶约束品性中假定的向量值函数，即 $\alpha(0)=x^*$、$d\alpha(0)/d\theta=z$(因为 $\hat{Z}^1(x^*)$ 是锥，不失一般性可假定 $\lambda=1$)，令 $d^2\alpha(0)/d\theta^2=w$. 从(3.94)和链式法则可得

$$\frac{d^2 g_i(\alpha(0))}{d\theta^2}=z^T\nabla^2 g_i(x^*)z+w^T\nabla g_i(x^*)=0, \quad i\in I(x^*), \quad (3.97)$$

$$\frac{d^2 h_j(\alpha(0))}{d\theta^2}=z^T\nabla^2 h_j(x^*)z+w^T\nabla h_j(x^*)=0, \quad j=1, \cdots, p. \quad (3.98)$$

从(3.72)至(3.74)和 $\hat{Z}^1(x^*)$ 的定义，我们有

$$\frac{df(\alpha(0))}{d\theta}=z^T\nabla f(x^*)=z^T\left[\sum_{i=1}^m \lambda_i^*\nabla g_i(x^*)+\sum_{j=1}^p \mu_j^*\nabla h_j(x^*)\right]=0. \quad (3.99)$$

因为 x^* 是局部极小值点而 $df(\alpha(0))/d\theta=0$，可知 $d^2 f(\alpha(0))/d\theta^2\geqslant 0$，即

$$\frac{d^2 f(\alpha(0))}{d\theta^2}=z^T\nabla^2 f(x^*)z+w^T\nabla f(x^*)\geqslant 0. \quad (3.100)$$

用相应的乘子乘(3.97)和(3.98)，从(3.100)减去，并利用(3.72)至(3.74)，我们得到

$$z^T\left[\nabla^2 f(x^*)-\sum_{i=1}^m \lambda_i^*\nabla^2 g_i(x^*)-\sum_{j=1}^p \mu_j^*\nabla^2 h_j(x^*)\right]z\geqslant 0. \quad (3.101)$$

∎

确认二阶约束品性至少和一阶的同样困难，然而有一种相对简单的情形，它蕴涵了一阶和二阶约束品性。如果向量 $\nabla g_i(x)$，$i\in I(x)$ 和 $\nabla h_j(x)$，$j=1, \cdots, p$ 是线性独立的，则两类约束品性在

$x \in X$ 都成立[28]。

例 **3.2.1**

McCormick[28]用下述问题去说明一种情况，即上述定理中的二阶必要条件能够用来减少满足一阶 Kuhn-Tucker 条件的点的数目*. 求参数 $\beta>0$ 的值，使得原点是下述问题的局部极小值点：

$$\min f(x_1, x_2) = (x_1-1)^2+(x_2)^2 \tag{3.102}$$

受限制于

$$g_1(x_1, x_2) = -x_1+\frac{(x_2)^2}{\beta} \geqslant 0. \tag{3.103}$$

一阶和二阶约束品性成立(为什么?), Lagrange 式给出为

$$L(x, \lambda) = (x_1-1)^2+(x_2)^2-\lambda\left[-x_1+\frac{(x_2)^2}{\beta}\right]. \tag{3.104}$$

对任何 $\beta\neq 0$, 取 $\lambda^*=2$, Kuhn-Tucker 条件在点 $x^*=(0, 0)$ 满足. 校核二阶必要条件，我们看到 $I(x^*)=\{1\}$, 且

$$\hat{Z}^1(x^*)=\{z: z\in R^2, z_1=0\}^{**}, \tag{3.105}$$

这样，$\hat{Z}^1(x^*)$ 中的任何向量有形式 $(0, z_2)^T$, 因此(3.96)这时变成

$$(0, z_2)\begin{bmatrix} 2 & 0 \\ 0 & 2-\dfrac{4}{\beta} \end{bmatrix}\begin{pmatrix} 0 \\ z_2 \end{pmatrix}=\left(2-\frac{4}{\beta}\right)(z_2)^2 \geqslant 0, \tag{3.106}$$

从而 $\beta<2$ 时，原点显然不是局部极小. 这样，满足 Kuhn-Tucker 条件的 β 值的集合大大地缩减了. ∎

现在让我们转向问题(P)中最优性的充分条件. 这样的条件由 Hestenes[21]、King[24]和 McCormick[28]导出，他们的结果后来被 Fiacco[14]加强.

用 $\hat{I}(x^*)$ 记这样的指标 i 的集合，对于它 $g_i(x^*)=0$ 且(3.72)至(3.74)为正数 λ_i^* 满足. 这样，$\hat{I}(x^*)$ 是 $I(x^*)$ 的子集. 令

$$\hat{\hat{Z}}^1(x^*)=\{z: z^T\nabla g_i(x^*)=0, i\in \hat{I}(x^*), z^T\nabla g_i(x^*)\geqslant 0, \\ i\in I(x^*), z^T\nabla h_j(x^*)=0, j=1, \cdots, p\}, \tag{3.107}$$

---

* 译注：指作为极小值点候选者的数目.

** 译注：原文为 $\{z: z\in R, z_1=0\}$，似漏写.

注意到 $\hat{\hat{Z}}^1(x^*)\subset Z^1(x^*)$，那么我们有下面的充分条件，它的证明是属于 McCormick 的[28]，这些条件是定理 2.8 的直接推广.

定理 3.11

设 x^* 对于问题(P)是能行的. 若存在向量 λ^* 和 μ^*，满足

$$\nabla_x L(x^*,\ \lambda^*,\ \mu^*)=\nabla f(x^*)-\sum_{i=1}^{m}\lambda_i^*\nabla g_i(x^*)-\sum_{j=1}^{p}\mu_j^*\nabla h_j(x^*)=0, \tag{3.108}$$

$$\lambda_i^* g_i(x^*)=0,\quad i=1,\ \cdots,\ m, \tag{3.109}$$

$$\lambda^*\geqslant 0, \tag{3.110}$$

且对每个 $z\neq 0$ 且 $z\in\hat{\hat{Z}}^1(x^*)$，可得

$$z^T\left[\nabla^2 f(x^*)-\sum_{i=1}^{m}\lambda_i^*\nabla^2 g_i(x^*)-\sum_{j=1}^{p}\mu_j^*\nabla^2 h_j(x^*)\right]z>0, \tag{3.111}$$

则 x^* 是问题(P)的严格局部极小值点.

【证明】 假定(3.108)至(3.111)成立而 x^* 不是严格局部极小值点，则存在一个收敛于 x^* 的能行点列 $\{z^k\}$，$z^k\neq x^*$，使得对每个 z^k，

$$f(x^*)\geqslant f(z^k). \tag{3.112}$$

令 $z^k=x^*+\theta^k y^k$，其中 $\theta^k>0$ 且 $\|y^k\|=1$. 不失一般性，假定序列$\{\theta^k,\ y^k\}$收敛到$(0,\ \bar{y})$，其中 $\|\bar{y}\|=1$. 因为点 z^k 是能行的，

$$g_i(z^k)-g_i(x^*)=\theta^k(y^k)^T\nabla g_i(x^*+\eta_i^k\theta^k y^k)\geqslant 0,\quad i\in I(x^*), \tag{3.113}$$

$$h_j(z^k)-h_j(x^*)=\theta^k(y^k)^T\nabla h_j(x^*+\bar{\eta}_j^k\theta^k y^k)=0,\quad j=1,\ \cdots,\ p, \tag{3.114}$$

而从(3.112)有

$$f(z^k)-f(x^*)=\theta^k(y^k)^T\nabla f(x^*+\eta^k\theta^k y^k)\leqslant 0, \tag{3.115}$$

其中 η^k，η_i^k，$\bar{\eta}_j^k$ 是 0 与 1 之间的数. 用 θ^k 除(3.113)至(3.115)并取极限，我们得到

$$(\bar{y})^T\nabla g_i(x^*)\geqslant 0,\quad i\in I(x^*), \tag{3.116}$$

$$(\bar{y})^T\nabla h_j(x^*)=0,\quad j=1,\ \cdots,\ p, \tag{3.117}$$

$$(\bar{y})^T\nabla f(x^*)\leqslant 0. \tag{3.118}$$

设对某个 $i\in\hat{I}(x^*)$，(3.116)作为严格的不等式成立，则联合(3.108)、(3.116)和(3.117)，我们得到

$$(\bar{y})^T\nabla f(x^*)=\sum_{i=1}^{m}\lambda_i^*(\bar{y})^T\nabla g_i(x^*)+\sum_{j=1}^{p}\mu_j^*(\bar{y})^T\nabla h_j(x^*)>0, \tag{3.119}$$

这与式(3.118)矛盾. 所以对一切 $i\in\hat{I}(x^*)$ 有 $(\bar{y})^T\nabla g_i(x^*)=0$，因而 $\bar{y}\in\hat{\hat{Z}}^1(x^*)$. 从 Taylor 定理我们得到

$$g_i(z^k)=g_i(x^*)+\theta^k(y^k)^T\nabla g_i(x^*)+\frac{1}{2}(\theta^k)^2(y^k)^T[\nabla^2 g_i(x^*+\xi_i^k\theta^k y^k)]y^k\geqslant 0,\quad i=1,\cdots,m, \tag{3.120}$$

$$h_j(z^k)=h_j(x^*)+\theta^k(y^k)^T\nabla h_j(x^*)+\frac{1}{2}(\theta^k)^2(y^k)^T[\nabla^2 h_j(x^*+\bar{\xi}_j^k\theta^k y^k)]y^k=0,\quad j=1,\cdots,p, \tag{3.121}$$

$$f(z^k)-f(x^*)=\theta^k(y^k)^T\nabla f(x^*)+\frac{1}{2}(\theta^k)^2(y^k)^T[\nabla^2 f(x^*+\xi^k\theta^k y^k)]y^k\leqslant 0, \tag{3.122}$$

其中 ξ^k, ξ_i^k, $\bar{\xi}_j^k$ 还是在0与1之间的数. 用相应的 λ_i^* 与 μ_j^* 乘(3.120)和(3.121)并从(3.122)中减去，就产生

$$\theta^k(y^k)^T\left\{\nabla f(x^*)-\sum_{i=1}^{m}\lambda_i^*\nabla g_i(x^*)-\sum_{j=1}^{p}\mu_j^*\nabla h_j(x^*)\right\}+\frac{1}{2}(\theta^k)^2(y^k)^T\Big[\nabla^2 f(x^*+\xi^k\theta^k y^k)-\sum_{i=1}^{m}\lambda_i^*\nabla^2 g_i(x^*+\xi_i^k\theta^k y^k)-\sum_{j=1}^{p}\mu_j^*\nabla^2 h_j(x^*+\bar{\xi}_j^k\theta^k y^k)\Big]y^k\leqslant 0. \tag{3.123}$$

由(3.108)，上述花括号中的表达式为0，用 $\frac{1}{2}(\theta^k)^2$ 除剩下的部分并取极限，我们得到

$$(\bar{y})^T\left[\nabla^2 f(x^*)-\sum_{i=1}^{m}\lambda_i^*\nabla^2 g_i(x^*)-\sum_{j=1}^{p}\mu_j^*\nabla^2 h_j(x^*)\right]\bar{y}\leqslant 0. \tag{3.124}$$

因为 $\bar{y}$ 非零且包含在 $\hat{\hat{Z}}^1(x^*)$ 中，可推知(3.124)与(3.111)矛盾. ▌

注意，出现在(3.111)方括号中的表达式是 $L(x, \lambda, \mu)$关于 x 的二阶导数矩阵. Fiacco 的著作[14]，将上面的定理扩充到不必是严格极小时的充分条件. 至于细节，有兴趣的读者可参考[14]和[15].

例 3.2.2

再考虑例 3.1.4 中给出的问题. 我们已经证明，(至少)有三个满足最优性必要条件的点. 让我们检验充分条件. 在 x^{*1}，有 $\hat{\hat{Z}}^1(x^{*1})=\{0\}$，从而没有向量 $z\neq 0$ 使 $z\in\hat{\hat{Z}}^1(x^{*1})$，所以定理 3.11 的充分条件自然满足. 读者可以验证这些条件在 x^{*2} 也成立. 但是在 x^{*3}，

$$\hat{\hat{Z}}^1(x^{*3})=\{(z_1, z_2): z_1=0\}, \tag{3.125}$$

而出现在(3.111)中的二次型是 $(-\sqrt{13})z^Tz$，它对于所有 $z\neq 0$ 是负的，这样 x^{*3} 不满足充分条件. ▌

3.3 Lagrange 式的鞍点

还有另一类型的最优性条件与 Lagrange 式有关且借助它的鞍点来表示. 这里我们叙述这些条件中的几个，对于出现于非线性规划问题(P)中的函数的性质，并不要求作特殊的假定. 有趣的是注意到，本章头两节中用到的可微性假定在某些未来的结果中可以舍去. 在更后几章中，我们将限于对特殊的函数类加强这些结果.

设 Φ 是两个向量变量 $x\in D\subset R^n$ 和 $y\in E\subset R^m$ 的实函数. Φ 的定义域就是 $D\times E$. 一个点 $(\bar{x}, \bar{y})$，$\bar{x}\in D$，$\bar{y}\in E$ 称为 Φ 的鞍点，如果对每个 $x\in D$ 和 $y\in E$ 有

$$\Phi(\bar{x}, y)\leqslant\Phi(\bar{x}, \bar{y})\leqslant\Phi(x, \bar{y}). \tag{3.126}$$

与非线性规划(P)相联系，有一个鞍点问题，它可叙述如下：

(S) 求一点 $\bar{x}\in R^n$，$\bar{\lambda}\in R^m$，$\bar{\lambda}\geqslant 0$ 和 $\bar{\mu}\in R^p$，使得 $(\bar{x}, \bar{\lambda}, \bar{\mu})$ 是 Lagrange 式

$$L(x, \lambda, \mu)=f(x)-\sum_{i=1}^{m}\lambda_i g_i(x)-\sum_{j=1}^{p}\mu_j h_j(x) \qquad (3.127)$$

的鞍点，即对一切 $x\in R^n$，$\lambda\in R^m$，$\lambda\geqslant 0$ 和 $\mu\in R^p$，有

$$L(\bar{x}, \lambda, \mu)\leqslant L(\bar{x}, \bar{\lambda}, \bar{\mu})\leqslant L(x, \bar{\lambda}, \bar{\mu}). \qquad (3.128)$$

在 Lagrange 式的鞍点与问题(P)的解之间的一个单侧关系在下面的结果中给出.

定理 3.12

若 $(\bar{x}, \bar{\lambda}, \bar{\mu})$ 是问题(S)的解，则 $\bar{x}$ 是问题(P)的解.

【证明】 设 $(\bar{x}, \bar{\lambda}, \bar{\mu})$ 是问题(S)的解，则对所有 $x\in R^n$，$\lambda\in R^m$，$\lambda\geqslant 0$ 和 $\mu\in R^p$，有

$$f(\bar{x})-\sum_{i=1}^{m}\lambda_i g_i(\bar{x})-\sum_{j=1}^{p}\mu_j h_j(\bar{x})\leqslant f(\bar{x})-\sum_{i=1}^{m}\bar{\lambda}_i g_i(\bar{x})-\sum_{j=1}^{p}\bar{\mu}_j h_j(\bar{x})$$
$$\leqslant f(x)-\sum_{i=1}^{m}\bar{\lambda}_i g_i(x)-\sum_{j=1}^{p}\bar{\mu}_j h_j(x). \qquad (3.129)$$

整理第一个不等式，我们得到

$$\sum_{i=1}^{m}(\bar{\lambda}_i-\lambda_i)g_i(\bar{x})+\sum_{j=1}^{p}(\bar{\mu}_j-\mu_j)h_j(\bar{x})\leqslant 0 \qquad (3.130)$$

对于所有 $\lambda\in R^m$，$\lambda\geqslant 0$ 和 $\mu\in R^p$ 成立. 现在假设对某个 k，$1\leqslant k\leqslant p$，有 $h_k(\bar{x})>0$. 对于 $i=1, \cdots, m$，令 $\lambda_i=\bar{\lambda}_i$，当 $j\neq k$ 时令 $\mu_j=\bar{\mu}_j$，令 $\mu_k=\bar{\mu}_k-1$，则假设 $h_k(\bar{x})>0$ 与(3.130)矛盾. 如果对某个 k，$h_k(\bar{x})<0$，我们可以选取适当的 μ，得到类似的矛盾. 这样，对 $j=1, \cdots, p$，$h_j(\bar{x})=0$. 现在令 $\mu=\bar{\mu}$ 和 $\lambda_1=\bar{\lambda}_1+1$，并对 $i=2, \cdots, m$，令 $\lambda_i=\bar{\lambda}_i$，我们得到 $g_1(\bar{x})\geqslant 0$. 令 $\lambda_2=\bar{\lambda}_2+1$，并对 $i=1, \cdots, m$，$i\neq 2$，令 $\lambda_i=\bar{\lambda}_i$，则 $g_2(\bar{x})\geqslant 0$. 对所有 i 重复这个过程，我们得到对 $i=1, \cdots, m$ 有 $g_i(\bar{x})\geqslant 0$. 由此可知 $\bar{x}$ 关于问题(P)是能行的. 然后令 $\lambda=0$，则由(3.129)中第一个不等式，我们有

$$0\leqslant -\sum_{i=1}^{m}\bar{\lambda}_i g_i(\bar{x}). \qquad (3.131)$$

但对 $i=1, \cdots, m$ 成立 $\bar{\lambda}_i\geqslant 0$ 和 $g_i(\bar{x})\geqslant 0$，所以，

$$\sum_{i=1}^{m}\bar{\lambda}_i g_i(\bar{x})=0, \qquad (3.132)$$

故对所有 i 成立 $\bar{\lambda}_i g_i(\bar{x})=0$.

现在转向(3.129)中的第二个不等式. 从上述论证可得

$$f(\bar{x})\leqslant f(x)-\sum_{i=1}^{m}\bar{\lambda}_i g_i(x)-\sum_{j=1}^{p}\bar{\mu}_j h_j(x). \tag{3.133}$$

如果 x 关于(P)是能行的, 则 $g_i(x)\geqslant 0$ 和 $h_j(x)=0$, 这样

$$f(\bar{x})\leqslant f(x), \tag{3.134}$$

$\bar{x}$ 就是(P)的解. 】

注意, 不论 Lagrange 式可微与否, 问题(P)中的上述最优性充分条件总成立. 然而, 如果函数 f, g_i, h_j 确是可微时, 我们有下述有趣的结果(与定理 3.2 比较之).

定理 3.13

设 f, g_1, $\cdots$, g_m, h_1, $\cdots$, h_p 是可微函数且 $(\bar{x}, \bar{\lambda}, \bar{\mu})$ 是问题(S)的解, 则 $Z^1(\bar{x})\cap Z^2(\bar{x})=\emptyset$, 且

$$\nabla_x L(\bar{x}, \bar{\lambda}, \bar{\mu})=\nabla f(\bar{x})-\sum_{i=1}^{m}\bar{\lambda}_i\nabla g_i(\bar{x})-\sum_{j=1}^{p}\bar{\mu}_j\nabla h_j(\bar{x})=0, \tag{3.135}$$

$$\bar{\lambda}_i g_i(\bar{x})=0, \quad i=1, \cdots, m, \tag{3.136}$$

$$\bar{\lambda}\geqslant 0. \tag{3.137}$$

【证明】 在前面定理的证明中我们看到, 如果 $(\bar{x}, \bar{\lambda}, \bar{\mu})$ 是(S)的解, 则 $\bar{\lambda}_i g_i(\bar{x})=0$, $i=1, \cdots, m$, 和 $\bar{\mu}_j h_j(\bar{x})=0$, $j=1, \cdots, p$. 显然按定义 $\bar{\lambda}$ 满足(3.137). 由(3.128)的第二个不等式可知, 对每个 $z\in R^n$ 和 $\alpha>0$ 有

$$L(\bar{x}, \bar{\lambda}, \bar{\mu})\leqslant L(\bar{x}+\alpha z, \bar{\lambda}, \bar{\mu}). \tag{3.138}$$

这样,

$$0\leqslant\frac{L(\bar{x}+\alpha z, \bar{\lambda}, \bar{\mu})-L(\bar{x}, \bar{\lambda}, \bar{\mu})}{\alpha}. \tag{3.139}$$

令 $\alpha\to 0$, 我们得到, 对每个 $z\in R^n$ 成立

$$0\leqslant z^T\nabla_x L(\bar{x}, \bar{\lambda}, \bar{\mu}). \tag{3.140}$$

所以, (3.135)必定成立. 由定理 3.2 可知, (3.135)至(3.137)成立的充要条件为 $Z^1(\bar{x})\cap Z^2(\bar{x})=\emptyset$. 】

注意, 虽然 Lagrange 式的鞍点蕴涵了(3.135)至(3.137)成立, 而无须附加的正则性条件, 但(P)的最优解 x^* 一般并不蕴涵满

足(3.135)至(3.137)的(λ^*, μ^*)的存在性，除非在(P)上加了象(3.71)这样的条件.

定理3.13和例3.1.1一起意味着，定理3.12的逆定理一般不成立．我们现在给出一个十分简单的例子以说明这点.

例 3.3.1

设我们有如下规划：

$$\min f(x)=x \tag{3.141}$$

受限制于

$$-(x)^2 \geqslant 0. \tag{3.142}$$

最优解是$x^*=0$．相应的Lagrange式的鞍点问题，是求$\lambda^* \geqslant 0$使得对每个$x \in R$有

$$x^*+\lambda(x^*)^2 \leqslant x^*+\lambda^*(x^*)^2 \leqslant x+\lambda^*(x)^2, \tag{3.143}$$

或者等价地，有

$$0 \leqslant x+\lambda^*(x)^2. \tag{3.144}$$

显然λ^*不能为0，但是对任何$\lambda^*>0$，我们可以选$x>-1/\lambda^*$，从而(3.144)将不成立．这样就不存在λ^*使得(x^*, λ^*)是鞍点． ∎

下一章将讨论这样的非线性规划，其中的目标函数与约束函数满足一定的凸性或凹性要求．如果这样的规划也满足某些约束品性，则定理3.12的逆定理对于它们成立．即(S)的解的存在性是(P)中最优性的必要条件.

鞍点是与对策论紧密相关的，其中利益冲突的两个局中人彼此反对对方．对于一个给定的“支付”函数$\Phi(x, y)$，一个局中人关于x求Φ的极小，而另一个局中人关于y求Φ的极大．这称为Φ的极小-极大化(或极大-极小化)．对策论及其在经济中应用的数学基础是由Von Neumann奠定的，在Von Neumann和Morgenstern的经典著作[80]中有描述.

我们建立鞍点与两个向量变量的函数的极小-极大化问题之间的联系来结束这一章.

引理 3.14

设Φ是两个向量变量$x \in D \subset R^n$和$y \in E \subset R^m$的实函数，则

$$\max_{y\in E}\min_{x\in D}\Phi(x, y)\leqslant\min_{x\in D}\max_{y\in E}\Phi(x, y), \tag{3.145}$$

只要上面的极小极大值与极大极小值存在.

【证明】 对任意固定的 $y\in E$, 我们有

$$\min_{x\in D}\Phi(x, y)\leqslant\Phi(x, y). \tag{3.146}$$

类似地, 对任意固定的 $x\in D$,

$$\Phi(x, y)\leqslant\max_{y\in E}\Phi(x, y). \tag{3.147}$$

因此对每个 $x\in D$ 和 $y\in E$,

$$\min_{x\in D}\Phi(x, y)\leqslant\max_{y\in E}\Phi(x, y). \tag{3.148}$$

结果得到

$$\max_{y\in E}\min_{x\in D}\Phi(x, y)\leqslant\min_{x\in D}\max_{y\in E}\Phi(x, y), \tag{3.149}$$

∎

例 3.3.2

一个实矩阵 $A=[a_{ij}]$ 可以看作是由

$$\Phi(i, j)=a_{ij} \tag{3.150}$$

给出的两变量 i, j 的实值函数. 令

$$A_1=\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \tag{3.151}$$

则

$$\max_j\min_i a_{ij}=\max(\min_i a_{i1}, \min_i a_{i2})=\max(-1, -1)=-1, \tag{3.152}$$

$$\min_i\max_j a_{ij}=\min(\max_j a_{1j}, \max_j a_{2j})=\min(1, 1)=1. \tag{3.153}$$

因此

$$\max_j\min_i\Phi(i, j)<\min_i\max_j\Phi(i, j), \tag{3.154}$$

且(3.145)作为严格不等式成立. 另一方面, 取

$$A_2=\begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}, \tag{3.155}$$

读者容易证明

$$\max_{j} \min_{i} a_{ij} = 2 = \min_{i} \max_{j} a_{ij}, \tag{3.156}$$

(3.145)作为方程成立. ▌

(3.145)中等号成立的必要和充分条件是Φ有鞍点. 形式地,有下列定理.

定理 3.15

设引理 3.14 的假设成立,则

$$\max_{y\in E} \min_{x\in D} \Phi(x, y) = \Phi(\bar{x}, \bar{y}) = \min_{x\in D} \max_{y\in E} \Phi(x, y) \tag{3.157}$$

的充要条件为$(\bar{x}, \bar{y})$是Φ的鞍点.

【证明】 设$(\bar{x}, \bar{y})$是Φ的鞍点,则

$$\max_{y\in E} \Phi(\bar{x}, y) \leqslant \Phi(\bar{x}, \bar{y}) \leqslant \min_{x\in D} \Phi(x, \bar{y}). \tag{3.158}$$

也有

$$\min_{x\in D} \max_{y\in E} \Phi(x, y) \leqslant \max_{y\in E} \Phi(\bar{x}, y), \tag{3.159}$$

$$\min_{x\in D} \Phi(x, \bar{y}) \leqslant \max_{y\in E} \min_{x\in D} \Phi(x, y). \tag{3.160}$$

联合上面三个关系式,我们得到

$$\min_{x\in D} \max_{y\in E} \Phi(x, y) \leqslant \Phi(\bar{x}, \bar{y}) \leqslant \max_{y\in E} \min_{x\in D} \Phi(x, y). \tag{3.161}$$

比较(3.161)和(3.145),我们断言(3.157)必定成立. 反之,设$(\bar{x}, y)$满足

$$\max_{y\in E} \Phi(\bar{x}, y) = \min_{x\in D} \max_{y\in E} \Phi(x, y) \tag{3.162}$$

和

$$\min_{x\in D} \Phi(x, \bar{y}) = \max_{y\in E} \min_{x\in D} \Phi(x, y), \tag{3.163}$$

若(3.157)成立,则由(3.162)、(3.163)以及极小和极大的定义,我们得到

$$\Phi(\bar{x}, y) \leqslant \max_{y\in E} \Phi(\bar{x}, y) = \Phi(\bar{x}, \bar{y}) = \min_{x\in D} \Phi(x, y) \leqslant \Phi(x, \bar{y}), \tag{3.164}$$

这样Φ有鞍点$(\bar{x}, \bar{y})$. ▌

练　习

3.A. 对于约束

$$(x_1)^2+(x_2-1)^2\leqslant 1, \tag{3.165}$$

$$x_2-x_1\leqslant 0, \tag{3.166}$$

求一个要极小化的目标函数 $f(x_1, x_2)$，使得在 $x^0=(0, 0)$ 有

$$Z^1(x^0)\cap Z^2(x^0)=\emptyset, \tag{3.167}$$

且这点不是一个局部最优点.

3.B. 设 A_1 和 A_2 是 R^n 的非空子集，证明 $A_1\subset A_2$ 蕴涵 $A_2'\subset A_1'$，即 A_2 的正法锥包含在 A_1 的正法锥中.

3.C. 证明对于问题(P)的能行集 X 中每一个 $\hat{x}$，$(Z^1(\hat{x}))'\subset(S(X, \hat{x}))'$.

3.D. 证明若 $\hat{x}$ 在集 $A\subset R^n$ 的内部，则 A 在 $\hat{x}$ 处的切锥是 R^n，并求 $(S(A, \hat{x}))'$.

3.E. 求超方体

$$A=\{x: x\in R^n,\ 0\leqslant x_j\leqslant 1,\ j=1,\ \cdots,\ n\} \tag{3.168}$$

在原点的切锥.

3.F. 证明引理 3.7 的下述推论：设 x^0 是问题(P)的解，且 x^0 在 X 的内部，则 $\nabla f(x^0)=0$.

3.G. 给定问题：

$$\max 4x_1+2x_2 \tag{3.169}$$

受限制于

$$x_1\log x_1-x_1+1\geqslant 0, \tag{3.170}$$

$$x_1\geqslant 0, \tag{3.171}$$

$$3x_1-2(x_2)^2\geqslant 1, \tag{3.172}$$

证明不存在点 $x^*\in R^2$ 满足最优性的 Kuhn-Tucker 必要条件．有没有满足定理 3.4 中 Fritz John 条件的点?

3.H. 为了求解问题：

$$\min(x_1)^2+(x_2)^2 \tag{3.173}$$

受限制于

$$(x_1-1)^3-(x_2)^2\geqslant 0, \tag{3.174}$$

建议借助于求相应的 Lagrange 式的逗留点．讨论在这样做时可能产生的困难.

3.I. 对于问题(P)的一种约束品性如下所述：向量 $\nabla g_i(x)$，$i\in I(x^*)$ 和 $\nabla h_j(x^*)$，$j=1, \cdots, p$，都是线性独立的．证明这个品性蕴涵 $S(X, x^*)=Z^1(x^*)$，但反之不真．［提示：令 $g(x)=(x_1)^2+(x_2)^2-x_2$，$h(x)=x_2$，

并取 $x^*=(0, 0, 0)$.]

3.J. 叙述关于下述问题的最优性的广义 Kuhn-Tucker 必要条件:

$$\max f(x) \tag{3.175}$$

受限制于

$$g_i(x)\geqslant 0, \quad i=1, \cdots, k, \tag{3.176}$$

$$g_i(x)\leqslant 0, \quad i=k+1, \cdots, m, \tag{3.177}$$

$$h_j(x)=0, \quad j=1, \cdots, p. \tag{3.178}$$

3.K. 证明下列属于 J. W. Gibbs(1876)的结果: 给定问题

$$\min f(x)=\sum_{j=1}^{n} f_j(x_j) \tag{3.179}$$

受限制于

$$x_j\geqslant 0, \quad j=1, \cdots, n, \tag{3.180}$$

$$\sum_{j=1}^{n} x_j=1, \tag{3.181}$$

其中 f 是可微函数. 令 x^* 是这问题的解, 则存在数 μ^* 使得

$$f'_j(x_j^*)=\mu^*, \quad x_j^*>0, \tag{3.182}$$

$$f'_j(x_j^*)\geqslant\mu^*, \quad x_j^*=0, \tag{3.183}$$

其中撇表示微商.

3.L. 给定非线性规划[37]

$$\min f(x)=\sum_{j=1}^{n} \frac{c_j}{x_j} \tag{3.184}$$

受限制于

$$\sum_{j=1}^{n} a_j x_j=b, \tag{3.185}$$

$$x\geqslant 0, \tag{3.186}$$

其中 a_j, b, c_j 是正常数. 证明目标函数的最优值由下式给出:

$$f(x^*)=\frac{\left[\sum_{j=1}^{n}(a_j c_j)^{1/2}\right]^2}{b}. \tag{3.187}$$

3.M. 对非线性规划:

$$\min f(x)=\sum_{j=1}^{n} c_j x_j \tag{3.188}$$

受限制于

$$g_i(x)\geqslant 0, \quad i=1, \cdots, m, \tag{3.189}$$

其中 c_j 是给定的不全为零的数, g_i 是可微函数. 证明若 x^* 是问题的解, (一阶)约束品性在 x^* 成立, 则 $I(x^*)\neq\emptyset$.

3.N. 在下面的问题中, 证明向量 $x^*=b/\|b\|$ 满足最优性的充分条件:

$$\max b^T x, \quad x \in R^n \tag{3.190}$$

受限制于

$$x^T x \leqslant 1, \tag{3.191}$$

其中向量 $b \neq 0$ 是给定的.

3.O. 给定约束

$$1-(x_1-2)^2-(x_2-5)^2 \geqslant 0, \tag{3.192}$$

$$1-(x_1-2)^2-\left(\frac{x_2}{4}\right)^2 \geqslant 0, \tag{3.193}$$

$$x_1-2 \geqslant 0, \tag{3.194}$$

画出能行集, 并证明对任何 $x \in X$,

$$(Z^1(x))' \neq (S(X, x))', \tag{3.195}$$

但二阶约束品性是满足的.

3.P. 非线性规划问题:

$$\max (x_2)^2-2x_1-(x_1)^2 \tag{3.196}$$

受限制于

$$(x_1)^2+(x_2)^2 \leqslant 1, \tag{3.197}$$

用某种数值算法在计算机上求解. 依赖于初始条件, 这算法收敛到四个不同的点:

(1) $x_1=1.000$, $x_2=0.000$;　　(2) $x_1=-1.000$, $x_2=0.000$;

(3) $x_1=-0.500$, $x_2=0.866$;　　(4) $x_1=-0.500$, $x_2=-0.866$;

(a) 检验所有这些点是否满足最优性的 Kuhn-Tucker 必要条件. 画出能行集并确定这些点上一阶约束品性是否成立.

(b) 假定二阶约束品性在前面每一点成立, 检验最优性的二阶必要条件.

(c) 最后, 导出在这问题中的严格最优性的充分条件, 并在满足必要条件的每个点上验证它们.

参 考 文 献

1. ABADIE, J., "On the Kuhn-Tucker Theorem," in *Nonlinear Programming*, J. Abadie (Ed.), North-Holland Publishing Co., Amsterdam, 1967.
2. ARROW, K. J., L. HURWICZ, and H. UZAWA, "Constraint Qualifications in Maximization Problems," *Naval Res. Log. Quart.*, **8**, 175-191 (1961).
3. BAZARAA, M. S., "Nonlinear Programming: Nondifferentiable Functions," Ph.D. dissertation, Georgia Institute of Technology, Atlanta, 1970.
4. BAZARAA, M. S., and J. J. GOODE, "Necessary Optimality Criteria in Mathematical Programming in the Presence of Differentiability," *J. Math. Anal. & Appl.*, 40, 609-621 (1972).

5. BAZARAA, M. S., J. J. GOODE, and C. M. SHETTY, "Constraint Qualifications Revisited," *Management Science*, **18**, 567–573 (1972).

6. BELTRAMI, E. J., "A Constructive Proof of the Kuhn-Tucker Multiplier Rule," *J. Math. Anal. & Appl.*, **26**, 297–306 (1967).

7. BERGTHALLER, C., private communication, 1971.

8. BRASWELL, R. N., and J. A. MARBAN, "Necessary and Sufficient Conditions for the Inequality Constrained Optimization Problem Using Directional Derivatives," *Int. J. Systems Sci.*, **3**, 263–275 (1972).

9. BRASWELL, R. N., and J. A. MARBAN, "On Necessary and Sufficient Conditions in Non-Linear Programming," *Int. J. Systems Sci.*, **3**, 277–286 (1972).

10. CANON, M. D., C. D. CULLUM, and E. POLAK, "Constrained Minimization Problems in Finite Dimensional Spaces," *SIAM J. Control*, **4**, 528–547 (1966).

11. DUBOVITSKII, A. Y., and A. A. MILYUTIN, "Extremum Problems in the Presence of Restrictions," *USSR Comp. Math. and Math Phys.*, **5**, 3, 1–80 (1965).

12. EVANS, J. P., "On Constraint Qualifications in Nonlinear Programming," *Naval Res. Log. Quart.*, **17**, 3, 281–286 (1970).

13. FARKAS, J., "Uber die Theorie der Einfachen Ungleichungen," *J. für die Reine und Angew. Math.*, **124**, 1–27 (1902).

14. FIACCO, A. V., "Second Order Sufficient Conditions for Weak and Strict Constrained Minima," *SIAM J. Appl. Math.*, **16**, 105–108 (1968).

15. FIACCO, A. V., and G. P. MCCORMICK, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley & Sons, New York, 1968.

16. GAMKRELIDZE, R. V., "Extremal Problems in Finite-Dimensional Spaces," *J. Optimization Theory and Appl.*, **1**, 173–193 (1967).

17. GOULD, F. J., and J. W. TOLLE, "A Necessary and Sufficient Qualification for Constrained Optimization," *SIAM J. Appl. Math.*, **20**, 164–172 (1971).

18. GOULD, F. J., and J. W. TOLLE, "Geometry of Optimality Conditions and Constraint Qualifications, *Math. Prog.*, **2**, 1–18 (1972).

19. GUIGNARD, M., "Generalized Kuhn-Tucker Conditions for Mathematical Programming Problems in a Banach Space," *SIAM J. Control*, **7**, 232–241 (1969).

20. HALKIN, H., and L. W. NEUSTADT, "General Necessary Conditions for Optimization Problems," *Proc. Natl. Acad. Sci.*, **56**, 1066–1071 (1966).

21. HESTENES, M. R., *Calculus of Variations and Optimal Control Theory*, John Wiley & Sons, New York, 1966.

22. JOHN, F., "Extremum Problems with Inequalities as Subsidiary Conditions," in *Studies and Essays, Courant Anniversary Volume*, K. D. Friedricks, et al. (Eds.), Interscience Publishers, New York, 1948.

23. KARLIN, S., *Mathematical Methods and Theory in Games, Programming and Economics*, Vols. I and II, Addison-Wesley Publishing Co., Reading, Mass., 1959.

24. KING, R. P., "Necessary and Sufficient Conditions for Inequality Constrained Extreme Values," *Ind. & Eng. Chem. Fund.*, **5**, 484–489 (1966).

25. KUHN, H. W., and A. W. TUCKER, "Nonlinear Programming," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman (Ed.), University of California Press, Berkeley, Calif., 1951.

26. MANGASARIAN, O. L., *Nonlinear Programming*, McGraw-Hill Book Co., New York, 1969.

27. MANGASARIAN, O. L., and S. FROMOVITZ, "The Fritz John Necessary Optimality Conditions in the Presence of Equality and Inequality Constraints," *J. Math. Anal. & Appl.*, **17**, 37–47 (1967).

28. MCCORMICK, G. P., "Second Order Conditions for Constrained Minima," *SIAM J. Appl. Math.*, **15**, 641–652 (1967).

29. MESSERLI, E. J., and E. POLAK, "On Second Order Necessary Conditions of Optimality," *SIAM J. Control*, **7**, 272–291 (1969).

30. NEUMANN, J. VON and O. MORGENSTERN, *Theory of Games and Economic Behavior*, 2nd ed., Princeton University Press, Princeton, N.J., 1947.

31. NEUSTADT, L. W., "An Abstract Variational Theory with Applications to a Broad Class of Optimization Problems. I. General Theory," *J. SIAM Control*, **4**, 505–527 (1966).

32. RITTER, K., "Optimization Theory in Linear Spaces. I," *Math. Ann.*, **182**, 189–206 (1969).

33. RITTER, K., "Optimization Theory in Linear Spaces Part II. On Systems of Linear Operator Inequalities in Partially Ordered Normed Linear Spaces," *Math. Ann.*, **183**, 169–180 (1969).

34. RITTER, K., "Optimization Theory in Linear Spaces Part III. Mathematical Programming in Partially Ordered Banach Spaces," *Math. Ann.*, **184**, 133–154 (1970).

35. SLATER, M., "Lagrange Multipliers Revisited: A Contribution to Nonlinear Programming," Cowles Commission Discussion Paper, Math. 403, November 1950.

36. VARAIYA, P., "Nonlinear Programming in Banach Space," *SIAM J. Appl. Math.*, **15**, 284–293 (1967).

37. WHITTLE, P., *Optimization Under Constraints*, Wiley-Interscience, London, 1971.

38. WILDE, D. J., "Differential Calculus in Nonlinear Programming," *Operations Research*, **10**, 764–773 (1962).

第 4 章

凸集和凸函数

在前一章开始部分引入的数学规划问题可算是我们将要处理的最一般的规划问题．事实上，对于如此一般的情形，已难以多加讨论．在本章中，我们将考察某些非线性规划，其中涉及的函数具有一些特殊的性质．以后将会明白，大多数这样的问题按某种方式联系着凸集和凸函数．本章提供一个基础，以理解这些问题的某些最重要的背景，诸如最优解的存在性、对偶性、最优解对小扰动的灵敏度及求解的数值算法．在这里，将概述与数学规划有关的凸集和凸函数的某些性质．这个概述无论如何不会是完备无缺的．从最优化方面的观点来看，对于凸集和凸函数的更完备的研究，读者可参考 Fenchel[8] 的讲义，或者 Rockafellar[13] 和 Stoer、Witzgall[15] 等新书．本章中出现的许多结果也能在 Berge[2] 中找到，那书透彻地阐明了基本的拓扑概念．

在六十年代后期，凸函数的概念已按许多途径被扩充，以推广关于凸函数的极值的结果，这些结果在数学规划中具有基本的重要性．这种推广将在以后几章中介绍．

4.1 凸　集

设 C 是 R^n 的子集，如果对任意两点 $x^1\in C$, $x^2\in C$, 连结它们的线段仍在 C 中，则称 C 为凸集．此处及以后三章中，q 常用来表示满足下列性质而称为权的有序数对 (q_1, q_2):

$$q_1\geqslant 0, \quad q_2\geqslant 0, \quad q_1+q_2=1. \tag{4.1}$$

这样，对集合 C，如果 $x^1\in C$, $x^2\in C$，而 q 如上所述，便蕴涵 $(q_1x^1+q_2x^2)\in C$，那么 C 就是凸集．

例 **4.1.1**

以下给出凸集的例子．

(a) 凸集的平凡的例子：空集，单个点 $x\in R^n$ 组成的集合，全空间 R^n。

(b) 设 $c\neq 0$ 是已给向量，b 是已给数，集合

$$H=\{x:x\in R^n,\ c^Tx=b\} \tag{4.2}$$

称为**超平面**. 它是一个凸集.

(c) 利用上述记号，我们定义**闭半空间**

$$H^c=\{x:x\in R^n,\ c^Tx\geqslant b\} \tag{4.3}$$

及**开半空间**

$$H^0=\{x:x\in R^n,\ c^Tx>b\}, \tag{4.4}$$

它们是由超平面 H 生成的. 注意每个超平面也产生反方向的闭半空间与开半空间，它们由倒转(4.3)、(4.4)中的不等号而得到. 所有这些半空间均为凸集.

(d) **超球**

$$S_\alpha(x^0)=\{x:x\in R^n,\ \|x-x^0\|\leqslant\alpha\}, \tag{4.5}$$

其中 $x^0\in R^n$ 是一个已给向量，α 是一个已给数. 可知超球是一个凸集. 】

现在给出凸集的某些简单的代数性质和几何性质.

假设已给 R^n 中的向量 $x^1,\ x^2,\ \cdots,\ x^p$，并已给满足下列条件的数 $\alpha_1,\ \alpha_2,\ \cdots,\ \alpha_p$:

$$\alpha_1\geqslant 0,\ \cdots,\ \alpha_p\geqslant 0,\qquad \sum_{i=1}^{p}\alpha_i=1, \tag{4.6}$$

那么向量 $x^1,\ x^2,\ \cdots,\ x^p$ 的**凸组合**由下式给出:

$$x=\alpha_1x^1+\alpha_2x^2+\cdots+\alpha_px^p. \tag{4.7}$$

定理 4.1

集合 $C\subset R^n$ 是凸集的充要条件为：点 $x^1\in C,\ \cdots,\ x^s\in C$ 的每个凸组合仍包含在 C 中.

【证明】 假设 C 为凸集. 下面对 s 用归纳法来证明. 当 $s=1$，定理显然正确. 现在设它对 $s\geqslant 1$ 正确，令

$$x=\alpha_1x^1+\cdots+\alpha_sx^s+\alpha_{s+1}x^{s+1}. \tag{4.8}$$

不失一般性，我们可以假定 $\alpha_{s+1}\neq 1$. 于是，能把上式写成

$$x=(1-\alpha_{s+1})z+\alpha_{s+1}x^{s+1}, \tag{4.9}$$

此处

$$z=\frac{\alpha_1}{1-\alpha_{s+1}}x^1+\cdots+\frac{\alpha_s}{1-\alpha_{s+1}}x^s. \tag{4.10}$$

由于

$$\frac{\alpha_1}{1-\alpha_{s+1}}\geqslant 0,\ \cdots,\ \frac{\alpha_s}{1-\alpha_{s+1}}\geqslant 0;\quad \sum_{i=1}^{s}\frac{\alpha_i}{1-\alpha_{s+1}}=1, \tag{4.11}$$

由归纳法假设，可知 $z\in C$，并由于 C 的凸性，就有 $x\in C$.

反之，如果 C 中的点的每个凸组合包含在 C 中，那么，特别地，$s=2$ 的结论是正确的；也就是说，$x^1\in C$、$x^2\in C$、$\alpha_1\geqslant 0$、$\alpha_2\geqslant 0$、$\alpha_1+\alpha_2=1$ 蕴涵着 $(\alpha_1x^1+\alpha_2x^2)\in C$，所以 C 是凸集. ■

定理 4.2

任意一组凸集之交集仍为凸集.

【证明】 设 x^1 和 x^2 是交集中的点，那么它们也包含在这组凸集中的任何一个集合之中，自然 $x=q_1x^1+q_2x^2$ 也如此，所以 x 也包含在交集中. ■

包含任意集合 $A\subset R^n$ 的所有凸集之交称为 A 的**凸包**，记为 $\mathrm{Co}(A)$. 由定理 4.2，显然集合 $\mathrm{Co}(A)$ 是凸集，它实际上是 R^n 中包含 A 的最小凸集. 两个集合 $X\subset R^n$，$Y\subset R^n$ 的和是指如下的集合：

$$X+Y=\{x+y: x\in X,\ y\in Y\}. \tag{4.12}$$

类似地，一个集合 $X\subset R^n$ 的数积定义为

$$\lambda X=\{\lambda x: x\in X,\ \lambda\in R\}. \tag{4.13}$$

定理 4.3

如果 C 和 D 是 R^n 中的凸集，λ 是一个实数，那么集合 $C+D$ 和 λC 也是凸集.

这个定理的证明留给读者.

在数学规划中，许多重要结果能够利用凸集的所谓**分离定理**来证明. 这些定理论述 R^n 中两个非空凸集 C_1 和 C_2，对于它们存在着超平面 H，使 C_1 落在 H 的一侧，而 C_2 落在另一侧. 这样的超平面称为分离 C_1 与 C_2 的超平面. 形式地说，设 S 和 T 是 R^n

的非空子集，如果 S 包含在由超平面 H 生成的一个闭半空间中，而 T 包含在另一闭半空间中，那么就说超平面 H **分离** S 和 T。这样的超平面称为**分离超平面**。如果 S 包含在由超平面 H 生成的一个开半空间中，而 T 包含在另一开半空间中，那么就说超平面 H **严格分离** S 和 T。读者从直观上可以清楚，给定了 R^n 中两个互不相交的凸集，可以找到分离它们的一个超平面。

现在我们叙述某些定理，它们论述这种分离超平面的存在性。这些定理能从几种途径来证明。我们的证明将利用拓扑的一些结果，特别是关于 R^n 中的紧致集的结果。关于这些定理的更完全的研究及有关结果，读者可以参见 Berge[2]。关于分离定理沿着其他线索的证明，例如能在 Rockafellar[13]中找到。

引理 4.4

设 C 是 R^n 中的非空闭凸集，它不包含原点，则存在一个超平面，它严格分离 C 和原点。

【证明】 令 $S_\alpha(0)$ 是以原点为中心的闭超球，即

$$S_\alpha(0)=\{x\in R^n:\|x\|\leqslant\alpha\},\tag{4.14}$$

且使 $C\cap S_\alpha(0)\neq\emptyset$。于是 $C\cap S_\alpha(0)$ 是一个紧致集，又连续函数 $\|x\|$ 在 $C\cap S_\alpha(0)$ 上达到极小，极小值点为某个 $x^0\in C$。

注意有 $\|x^0\|>0$。于是，对每个 $x\in S_\alpha(0)\cap C$ 有 $\|x\|\geqslant\|x^0\|$，同时，按 $S_\alpha(0)$ 的定义，对每个 $x\in C$ 也成立 $\|x\|\geqslant\|x^0\|$。下面设 x 是 C 中任一点。因为 C 是凸集，对任何 $1\geqslant\lambda\geqslant0$，我们有 $(\lambda x+(1-\lambda)x^0)\in C$，且

$$\|\lambda x+(1-\lambda)x^0\|^2\geqslant\|x^0\|^2,\tag{4.15}$$

所以有

$$(\lambda x+(1-\lambda)x^0)^T(\lambda x+(1-\lambda)x^0)\geqslant(x^0)^Tx^0,\quad 1\geqslant\lambda\geqslant0,\tag{4.16}$$

也就是

$$(\lambda)^2(x-x^0)^T(x-x^0)+2\lambda(x^0)^T(x-x^0)\geqslant0,\quad 1\geqslant\lambda\geqslant0.\tag{4.17}$$

假如 $(x^0)^T(x-x^0)=-\varepsilon<0$，我们便能选择充分小的 λ 使

$$2\varepsilon > \lambda(x-x^0)^T(x-x^0) > 0. \tag{4.18}$$

于是便有

$$2\lambda(x^0)^T(x-x^0) < -(\lambda)^2(x-x^0)^T(x-x^0), \tag{4.19}$$

从而与(4.17)矛盾，因此我们必定有

$$(x^0)^T(x-x^0) \geqslant 0, \tag{4.20}$$

或者说，对每个 $x\in C$ 均成立

$$(x^0)^T x \geqslant (x^0)^T x^0 > 0. \tag{4.21}$$

令 $\alpha=(x^0)^T(x^0)/2$，于是超平面

$$H_C=\{x: x\in R^n,\ (x^0)^T x=\alpha\} \tag{4.22}$$

严格分离 C 与原点. 】

现在，我们能够证明以下的定理.

定理 4.5(严格分离定理)

设 C_1 和 C_2 是 R^n 中两个互不相交的非空闭凸集，并设 C_2 是紧致的，则存在一个超平面严格分离 C_1 和 C_2。

【证明】 由于 C_2 是紧致集，所以 $C_1-C_2=C_1+(-C_2)$ 是闭集(例如参见[1, 2, 3])，并且由定理 4.3，它是凸集. 由于 C_1 和 C_2 互不相交，C_1-C_2 便不包含原点. 根据引理 4.4，必存在一个超平面

$$H_{C_1-C_2}=\{x: x\in R^n,\ (x^0)^T x=\alpha\}, \tag{4.23}$$

它严格分离 C_1-C_2 和原点，此处 $x^0\in C_1-C_2$ 使 C_1-C_2 到原点的距离达到极小，且 $\alpha=(x^0)^T(x^0)/2$.

对每个 $x\in C_1-C_2$，有

$$(x^0)^T x > \alpha > 0. \tag{4.24}$$

令 $x=u-v$，这里 $u\in C_1$，$v\in C_2$，我们得到

$$(u^0-v^0)^T(u-v) > \alpha > 0, \tag{4.25}$$

所以

$$\inf_{u\in C_1}(u^0-v^0)^T u \geqslant \sup_{v\in C_2}(u^0-v^0)^T v+\alpha > \sup_{v\in C_2}(u^0-v^0)^T v. \tag{4.26}$$

选择一数 β 使

$$\inf_{u\in C_1}(u^0-v^0)^T u > \beta > \sup_{v\in C_2}(u^0-v^0)^T v, \tag{4.27}$$

于是超平面$\{x: x\in R^n, (x^0)^T x=\beta\}$严格分离$C_1$和$C_2$. 】

在关于严格分离两个凸集的上述结果中，关键性的假定是闭集条件. 如果我们不限定它们为闭凸集，那还可以得到类似于引理 4.4 和定理 4.5 的某种较弱的结论.

引理 4.6

设C是R^n中非空凸集，它不包含原点，则存在一个超平面分离C和原点.

【证明】 对每个$x\in C$，令

$$Y(x)=\{y: y\in R^n, y^T y=1, y^T x\geqslant 0\}, \tag{4.28}$$

这个集合是非空闭集.

设$x^1, \cdots, x^k$是C中任意有限点集. 因为C为凸集，根据定理 4.1，下式给出的x属于C:

$$x=\alpha_1 x^1+\cdots+\alpha_k x^k, \quad \sum_{i=1}^{k}\alpha_i=1, \quad \alpha_i\geqslant 0, \tag{4.29}$$

这些x的集合是C中紧致集(它是$x^1, \cdots, x^k$组成的点集的凸包). 由引理 4.4，必存在$y^0\neq 0$，使

$$(y^0)^T x^i>0, \quad i=1, \cdots, k. \tag{4.30}$$

不失一般性，我们可以假定上述y^0满足$(y^0)^T y^0=1$，于是

$$\bigcap_{i=1}^{k} Y(x^i)\neq\emptyset. \tag{4.31}$$

由于$Y(x)$是紧致集$Y=\{y: y\in R^n, y^T y=1\}$的闭子集，根据有限交集性质[2, 12]可知

$$\bigcap_{x\in C} Y(x)\neq\emptyset. \tag{4.32}$$

选取任一$\hat{y}\in\bigcap_{x\in C}Y(x)$，于是对每个$x\in C$均有$\hat{y}^T x\geqslant 0$，所以超平面$\{x: x\in R^n, \hat{y}^T x=0\}$分离$C$和原点. 】

例 4.1.2

考察集合

$$C=\{x: x\in R^2, x_1>0\}. \tag{4.33}$$

这集合是一个非空开凸集，它不包含原点. 超平面$\{x: x\in R^2, x_1=0\}$分离C和原点. 然而，不存在超平面严格分离它们. 】

定理 4.7(分离定理)

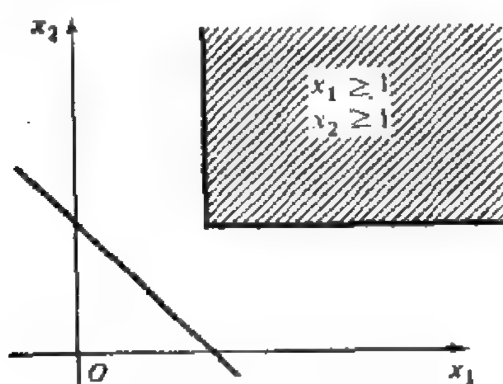
设 C_1 和 C_2 是 R^n 中两个互不相交的非空凸集. 则存在一个超平面分离它们.

【证明】 集合 C_1-C_2 是凸集, 且不包含原点. 根据引理 4.6, 存在一个向量 $\hat{y}$, 使对每个 $x\in C_1-C_2$, 均有 $\hat{y}^Tx\geqslant 0$, 或者, 等价地, 对 $u\in C_1$, $v\in C_2$, 必有 $\hat{y}^T(u-v)\geqslant 0$. 所以存在一个数 β 满足

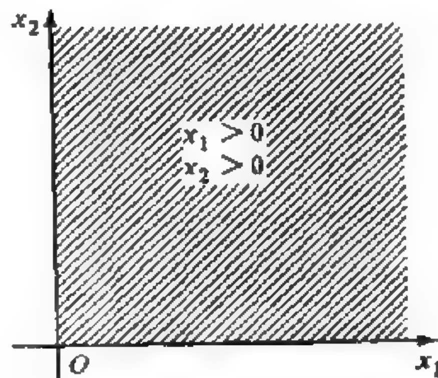
$$\inf_{u\in C_1}\hat{y}^Tu\geqslant\beta\geqslant\sup_{v\in C_2}\hat{y}^Tv, \tag{4.34}$$

于是超平面 $\{x:x\in R^n,\ \hat{y}^Tx=\beta\}$ 分离 C_1 和 C_2. ∎

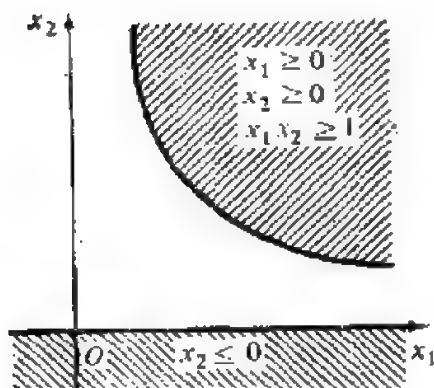
图 4.1 提供关于凸集的分离定理的补充说明. 在 (a) 中是一个严格分离的简单情形, 而 (b) 显示了不包含原点的开凸集并非总能将它与原点严格分离. (b) 中的例子说明, 引理 4.4 中闭凸集的



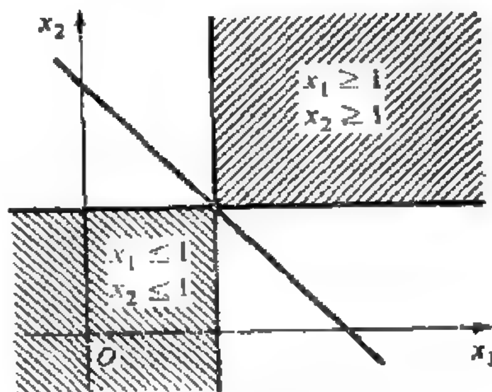
(a) 严格分离一个闭凸集与原点的超平面



(b) 不能和原点严格分离的开凸集



(c) 不能严格分离的两个非紧致闭凸集



(d) 对于具有公共边界点的凸集的分离超平面

图 4.1 凸集的分离

假设是决定性的．类似地，正如(c)中的集合所表明的，定理 4.5 中 C_2 为紧致的假设是重要的．最后，在(d)中我们对定理 4.7 的假设条件给出了一个可能的减弱．关于 C_1 与 C_2 互不相交的假设，可以代之以 C_1 和 C_2 没有公共内点的假设，这样，它们可以有公共的边界点．

作为上述分离定理的第一个应用，我们现在证明著名的 Farkas 引理，该引理已在第 3 章引入，为了方便，现在重录于下．

引理 4.8(Farkas 引理)

假设 A 为给定的 $m\times n$ 实矩阵，b 为给定的 n 维向量，那么，对所有满足 $Ay\geqslant 0$ 的向量 y 均成立 $b^Ty\geqslant 0$ 的充要条件为：存在 m 维向量 $\rho\geqslant 0$ 使 $A^T\rho=b$．

【证明】 这个结论等价于说，$Ay\geqslant 0$，$b^Ty<0$ 有解的充要条件是 $A^T\rho=b$ 与 $\rho\geqslant 0$ 无解．而后一组无解的充要条件是下列非空闭凸集互不相交(参见习题 4.C)：

$$C_1=\{x:x\in R^n,\ x=A^T\rho,\ \rho\geqslant 0\},\ C_2=\{b\},\tag{4.35}$$

注意 C_2 是紧致的．

如果上述条件成立，则由定理 4.5，存在一个非零向量 $c\in R^n$ 及一实数 α 使 $c^Tb<\alpha$．并对每个 $x\in C_1$ 有 $c^Tx>\alpha$，或者等价地，对每个 $\rho\geqslant 0$ 有 $c^TA^T\rho>\alpha$．

令 $\rho=0$，我们得到 $\alpha<0$．又对 $j=1, 2, \cdots, m$，令 $\rho=(0,\cdots,\rho_j, 0, \cdots, 0)$，其中 ρ_j 是一个很大的正数，我们得到 $c^TA^T\geqslant 0$，因此 $Ac\geqslant 0$．所以 c 是 $Ay\geqslant 0$ 与 $b^Ty<0$ 的解．反之，若 $A^T\rho=b$，$\rho\geqslant 0$ 又 $Ay\geqslant 0$，则 $b^Ty=\rho^TAy\geqslant 0$． ▌

分离定理还能用来证明关于其他线性方程与不等式组的类似结果，这些结果的广泛内容可在 Mangasarian[12]中找到．除数学规划的目的外，凸集还有一定的研究兴趣，可参见 Eggleston[5]、Fenchel[8]、Rockafellar[13] 或 Valentine[16]．

4.2 凸 函 数

在进入本世纪前后，Hölder[9]和 Jensen[10]奠定了凸函数理论

的基础．从几何观点来观察凸函数是直观的，在大多数场合也是足够的．假如曲线 $f(x)$ 上任意两点连成的弦落在曲线上或曲线的上侧，那么实值函数 f 称为凸函数．图 4.2 所表示的凸函数说明了"弦在曲线以上"的原则．观察一下这些图形还可看到，凸函数在其定义域中并非必然是到处连续或可微的．另一个重要的观察是凸集和凸函数之间的密切联系，它将作为我们对凸函数进行形式研究的出发点．图 4.2 中这些曲线上方的阴影区域都是凸集，这同样也能用来定义凸函数．由于以后将会明白的原因，允许凸函数取值无穷大将是方便的．

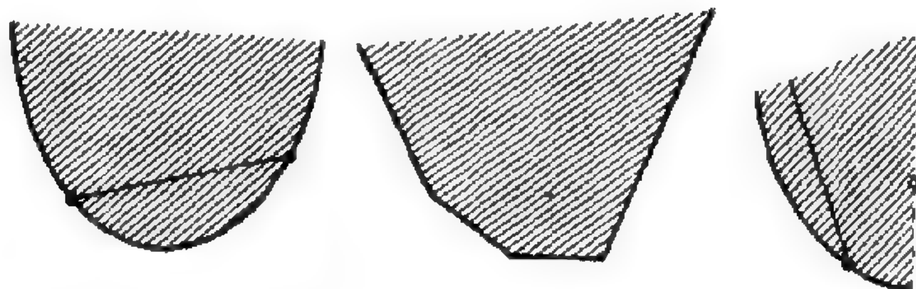


图 4.2　凸函数的例子

设 f 是定义在 R^n 的子集 D 上的函数，取值在扩充的实数集上，亦即，$f(x)$ 或是实数或是 $\pm\infty$．R^{n+1} 中的子集

$$P(f)=\{(x,\ \alpha):x\in D\subset R^n,\ \alpha\in R,\ f(x)\leqslant\alpha\} \tag{4.36}$$

称为 f 的**上图象**．如果 $P(f)$ 为凸集，我们定义 f 为**凸函数**．以下是一些直接明了的例子．

例 4.2.1

R^n 上函数 $f=+\infty$ 是一个凸函数，因为 $P(f)=\emptyset$，而空集是凸集．类似地，R^n 上 $f=-\infty$ 也是凸函数，因为 $P(f)=R^{n+1}$．

考虑定义在 R^n 的真子集 D 上的凸函数 f．令

$$f_1(x)=\begin{cases} f(x), & x\in D, \\ +\infty, & x\overline{\in} D. \end{cases} \tag{4.37}$$

定义在 D 上的函数 f 的上图象 $P(f)$ 等价于 $P(f_1)$，其中 f_1 是定义在整个 R^n 上的函数．用这种办法，我们总能构造定义在整个 R^n 上的凸函数．特别地，设 $a\in R$、$b\in R^n$ 是已给的，则下列函数

是定义在 R^n 上的凸函数：

$$f_1(x)=\begin{cases}a, & x=b,\\ +\infty, & x\neq b.\end{cases} \tag{4.38}$$

∎

作为例 4.2.1 的一个结果，以后除明确地提到的之外，我们总认为凸函数定义在整个 R^n 上．集合

$$\mathrm{ED}(f)=\{x: x\in R^n,\ f(x)<+\infty\} \tag{4.39}$$

称为 f 的**有效区域**，或**有限性区域**，实际上它是 $P(f)$ 在 R^n 上的投影．如 f 是一个凸函数，则 $\mathrm{ED}(f)$ 是 R^n 中的凸集．相反的结论一般不成立，这就是说，如果 $\mathrm{ED}(f)$ 是凸集，那么 f 不必一定是凸函数．

设一个函数定义在 $D\subset R^n$ 上，如果改变符号后它成为凸函数，则称它为**凹函数**．换言之，当且仅当 $-g$ 为凸函数时，g 为凹函数．集合

$$Q(g)=\{(x,\ \alpha): x\in D\subset R^n,\ \alpha\in R,\ g(x)\geqslant\alpha\} \tag{4.40}$$

称为 g 的**下图象**，而当 $Q(g)$ 为 R^{n+1} 中凸集时，称 g 为凹函数．以下研究中，在多数情形，我们将只处理凸函数．读者应当能够修改关于凸函数性质的结论，以得到凹函数的类似性质．

允许凸函数取值无穷大，便需要在算术运算中保持一定的谨慎，例如不能进行不确定的运算 $+\infty+(-\infty)$．具有值 $+\infty$ 与 $-\infty$ 的凸函数在应用中不常出现．因此，我们将主要关心**正常凸函数**，它定义为无处取值 $-\infty$ 且不恒等于 $+\infty$ 的凸函数．不是正常的凸函数就称为**非正常的**．形式地说，若 f 是凸函数，对一切 $x\in R^n$ 有 $f(x)>-\infty$，且 $\mathrm{ED}(f)\neq\emptyset$，则 f 是正常凸函数．对于 $\mathrm{ED}(f)$ 多于一点的正常凸函数，我们现在能推导一个更为熟悉的结果，它可作为另一定义．已知 $x^1\in\mathrm{ED}(f)$，$x^2\in\mathrm{ED}(f)$，如 $(x^1,\ \alpha^1)\in P(f)$，$(x^2,\ \alpha^2)\in P(f)$，那么对任何权向量 q，我们有 $(q_1x^1+q_2x^2,\ q_1\alpha^1+q_2\alpha^2)\in P(f)$．所以

$$f(q_1x^1+q_2x^2)\leqslant q_1\alpha^1+q_2\alpha^2. \tag{4.41}$$

对每个适合 $f(x^1)\leqslant\alpha^1$，$f(x^2)\leqslant\alpha^2$ 的 α^1、α^2，不等式(4.41)必须成

立，所以(4.41)可以修改为

$$f(q_1x^1+q_2x^2)\leqslant q_1f(x^1)+q_2f(x^2). \tag{4.42}$$

不等式(4.42)便是实的凸函数的“经典”定义，它表达了“弦在曲线上方”的性质．假如我们采用算术规则 $0\cdot(\infty)=0\cdot(-\infty)=0$，并避免不确定的运算 $+\infty+(-\infty)$，那么即使对某些 x 有 $f(x)=+\infty$ 或 $f(x)=-\infty$，(4.42)仍一般地保持成立．

设实值函数 f 定义在 R^n 的凸子集 C 上，如果对任何 $x^1\in C$、$x^2\in C$，$x^1\neq x^2$ 及 $1>q_1>0$，$q_1+q_2=1$，不等式(4.42)成立严格的不等号，那末 f 称为**严格凸函数**．对于严格凸函数，本章所述有关凸函数的许多结果还可加强，这请读者去做．我们仅在一、二个结果中，特别地提及严格凸函数．

下面让我们考查凸函数的一些基本性质．

定理 4.9

设 f 是 R^n 上的正常凸函数．令 $x^1, \cdots, x^s$ 是 R^n 中的点，且 $q_1, \cdots, q_s$ 是非负数，满足 $q_1+\cdots+q_s=1$．则

$$f(q_1x^1+\cdots+q_sx^s)\leqslant q_1f(x^1)+\cdots+q_sf(x^s). \tag{4.43}$$

【证明】 如对某个 i 有 $f(x^i)=+\infty$，那么(4.43)就平凡地成立．现在设对所有 i 均有 $f(x^i)<+\infty$．f 的上图象是一个凸集，由定理 4.1，它必包含它的一些点的每个凸组合．所以 $(x^1, \alpha^1)\in P(f), \cdots, (x^s, \alpha^s)\in P(f)$ 必蕴涵 $(q_1x^1+\cdots+q_sx^s,\ q_1\alpha^1+\cdots+q_s\alpha^s)\in P(f)$，此即

$$f(q_1x^1+\cdots+q_sx^s)\leqslant q_1\alpha^1+\cdots+q_s\alpha^s. \tag{4.44}$$

因为 α^i 只须满足 $f(x^i)\leqslant\alpha^i$，$i=1, \cdots, s$，所以不等式(4.43)成立． ▌

定理 4.10

设 f 是凸函数，λ 是非负数，则 λf 也是凸函数．

设 f 和 g 是凸函数，则 $f+g$ 也是凸函数，这里假定不会出现不确定的运算 $+\infty+(-\infty)$．

证明留给读者．由此定理立即可得以下结论．

推论 4.11

在定理 4.10 的假定下，设 $\lambda_1 \geqslant 0, \cdots, \lambda_k \geqslant 0$，则凸函数的线性组合 $\lambda_1 f_1 + \cdots + \lambda_k f_k$ 也是凸函数.

设函数 Ψ 定义在 R 上，取值在扩充的实数集中，如果对任何 $x^1 < x^2$，均有 $\Psi(x^1) \leqslant \Psi(x^2)$，则称 Ψ 为非减函数. 下列定理时常有助于判别凸函数，或者从已有的凸函数去构造新的凸函数.

定理 4.12

设 f 是 R^n 上的实值凸函数，且设 Ψ 是 R 上的非减正常凸函数，则 $\Psi(f(x))$ 是 R^n 上的凸函数.

【证明】 由(4.42)，对每个 q 我们有

$$f(q_1 x^1 + q_2 x^2) \leqslant q_1 f(x^1) + q_2 f(x^2). \tag{4.45}$$

因为 Ψ 是非减的，得

$$\Psi(f(q_1 x^1 + q_2 x^2)) \leqslant \Psi(q_1 f(x^1) + q_2 f(x^2)), \tag{4.46}$$

再由 Ψ 的凸性，所以

$$\Psi(f(q_1 x^1 + q_2 x^2)) \leqslant q_1 \Psi(f(x^1)) + q_2 \Psi(f(x^2)). \tag{4.47}$$

■

注意，通过规定 $\Psi(+\infty) = +\infty$，定理 4.12 可以扩充到 R^n 上的正常凸函数 f.

图 4.2 中出现的函数之一是分段线性函数. 注意到线性函数是凸函数，并利用下列定理，可以证明这个分段线性函数的凸性.

定理 4.13

设 $\{f_i\}$，$i \in I$，表示有限个或无限个 R^n 上凸函数的总体，对每个 $x \in R^n$，定义这个总体的点式上确界如下：

$$f(x) = \sup_{i \in I} f_i(x), \tag{4.48}$$

则 f 是一个凸函数.

【证明】 f_i 的上图象都是凸集，由定理 4.2，所有 $P(f_i)$ 的交也是凸集. 根据定义，有

$$\bigcap_{i \in I} P(f_i) = \{(x, \alpha) : x \in R^n,\ \alpha \in R,\ \forall i \in I \text{ 有 } f_i(x) \leqslant \alpha\}, \tag{4.49}$$

这样我们得到

$$\bigcap_{i\in I} P(f_i)=\{(x,\ \alpha):x\in R^n,\ \alpha\in R,\ \sup_{i\in I} f_i(x)=f(x)\leqslant\alpha\},\tag{4.50}$$

所以 f 是一个凸函数. ▌

我们已经看到, R^n 上每个凸函数 f 能够联系 R^{n+1} 中的一个凸集 $P(f)$. 下一结果论述逆命题, 即对 R^{n+1} 中的凸集, 通过取它的"下部轮廓", 可以构造出 R^n 上的凸函数. 注意, 在空集上取下确界规定为 $+\infty$.

定理 4.14

设 C 是 R^{n+1} 中的一个凸集, 对每个 $x\in R^n$, 令 f 定义为

$$f(x)=\inf\{\alpha:\alpha\in R,\ (x,\ \alpha)\in C\},\tag{4.51}$$

则 f 是 R^n 上的凸函数.

【证明】 我们必须证明 $P(f)$ 是凸集. 设 $(x^1,\ \alpha^1)$ 和 $(x^2,\ \alpha^2)$ 是 $P(f)$ 中任意两点. 由 f 的定义可知, 必存在 $\hat{\alpha}^1\leqslant\alpha^1$ 及 $\hat{\alpha}^2\leqslant\alpha^2$ 使 $(x^1,\ \hat{\alpha}^1)\in C$ 及 $(x^2,\ \hat{\alpha}^2)\in C$. 因为 C 是凸集, 由 (4.51), 对每个 q 我们有

$$f(q_1x^1+q_2x^2)\leqslant q_1\hat{\alpha}^1+q_2\hat{\alpha}^2\leqslant q_1\alpha^1+q_2\alpha^2.\tag{4.52}$$

所以 $(q_1x^1+q_2x^2,\ q_1\alpha^1+q_2\alpha^2)\in P(f)$. ▌

例 4.2.2

设 C 是开单位圆, 即

$$C=\{(x_1,\ x_2):(x_1)^2+(x_2)^2<1\},\tag{4.53}$$

这时, 由 (4.51) 构造的 R 上的凸函数 f 可由下式表示(见图 4.3):

$$f(x)=\begin{cases}-[1-(x)^2]^{1/2}, & |x|<1,\\ +\infty, & |x|\geqslant 1.\end{cases}\tag{4.54}$$ ▌

下一个定理说明, 在凸集 C 上的函数为凸函数的充要条件是: 当 f 限制在 C 中每个线段上时为凸函数.

定理 4.15

f 是 R^n 上凸函数的充要条件为: 对任何 $x^1\in R^n$、$x^2\in R^n$, 由下式定义的函数 ϕ 对整个 $0\leqslant\lambda\leqslant 1$ 为凸函数:

$$\phi(\lambda)=f(\lambda x^1+(1-\lambda)x^2).\tag{4.55}$$

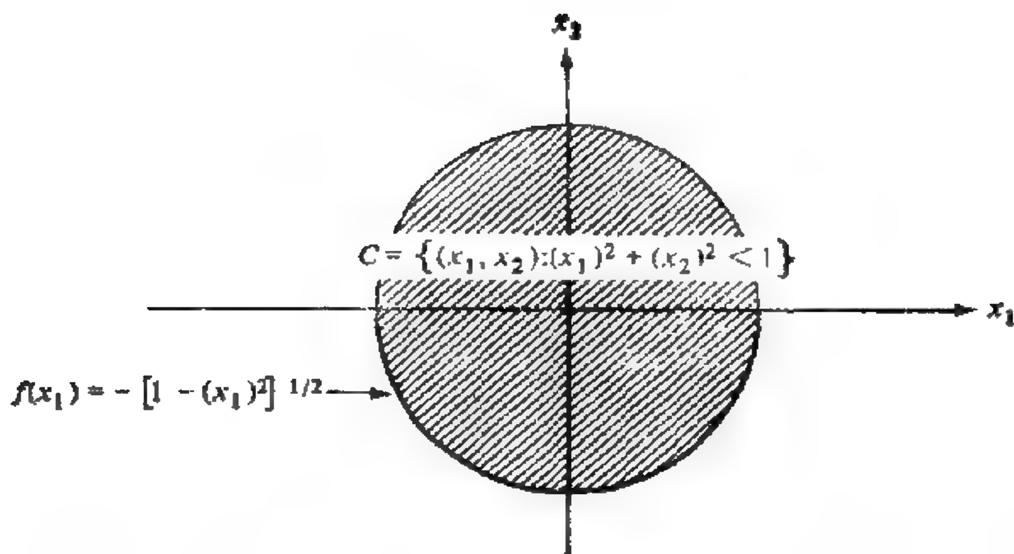


图 4.3 开单位圆的下部轮廓

【证明】 设 f 是 R^n 上的凸函数，并设 x^1、x^2 是 R^n 中任意点．ϕ 的上图象由下式给出：

$$P(\phi)=\{(\lambda, \alpha): \lambda \in R,\ 0 \leqslant \lambda \leqslant 1,\ \alpha \in R,\ \phi(\lambda) \leqslant \alpha\}, \tag{4.56}$$

我们必须证明 $P(\phi)$ 是一个凸集．设 $(\lambda^1, \alpha^1) \in P(\phi)$，$(\lambda^2, \alpha^2) \in P(\phi)$，且令

$$z^1=\lambda^1 x^1+(1-\lambda^1) x^2,\ z^2=\lambda^2 x^1+(1-\lambda^2) x^2. \tag{4.57}$$

那么我们有

$$f(z^1)=\phi(\lambda^1) \leqslant \alpha^1, \tag{4.58}$$

$$f(z^2)=\phi(\lambda^2) \leqslant \alpha^2, \tag{4.59}$$

因此 $(z^1, \alpha^1) \in P(f)$，$(z^2, \alpha^2) \in P(f)$；又由于 $P(f)$ 为凸集，对每个 q 我们也有 $(q_1 z^1+q_2 z^2,\ q_1 \alpha^1+q_2 \alpha^2) \in P(f)$．

因此

$$f(q_1 z^1+q_2 z^2) \leqslant q_1 \alpha^1+q_2 \alpha^2. \tag{4.60}$$

由(4.55)和(4.57)，我们得到

$$f(q_1 z^1+q_2 z^2)=\phi(q_1 \lambda^1+q_2 \lambda^2), \tag{4.61}$$

所以 $(q_1 \lambda^1+q_2 \lambda^2,\ q_1 \alpha^1+q_2 \alpha^2) \in P(\phi)$，这就是说，$\phi$ 为凸函数．逆命题的证明是类似的，留给读者． ▌

读者可能已注意到，我们还没有讲到凸函数的连续性．容易

说明，凸函数可能是不连续的．例如，取函数 f 为(见图 4.4)：

$$f(x)=\begin{cases}(x)^2-1, & x<1,\\ 2, & x=1,\\ +\infty, & x>1,\end{cases} \tag{4.62}$$

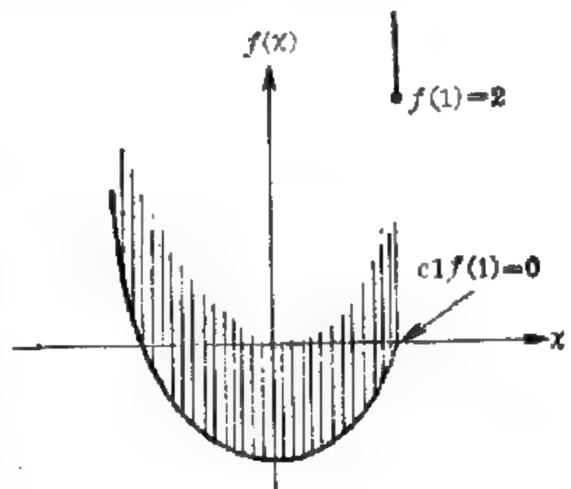


图 4.4　凸函数的闭包运算

这是一个凸函数，在它的有效区域的边界上是不连续的．粗略地说，凸函数的间断性只能在其有效区域的边界点上出现，利用凸函数的闭包运算，可以部分地消除这种不连续性．考虑定义在 R^n 上的凸函数 f，取形如 $h(x)=a^Tx-b$ 的函数，要求对任意 $x\in R^n$ 均成立 $h(x)\leqslant f(x)$，* 并考虑所有这种线性仿射函数的总体．我们定义支撑集如下：

$$L(f)=\{(a,\ b): a\in R^n,\ b\in R,\ \forall x\in R^n \text{ 有 } a^Tx-b\leqslant f(x)\}. \tag{4.63}$$

然后我们定义凸函数的闭包，记作 $\mathrm{cl}\,f$，为

$$\mathrm{cl}\,f(x)=\sup_{(a,b)\in L(f)}\{a^Tx-b\}. \tag{4.64}$$

显然，对一切 $x\in R^n$，成立 $\mathrm{cl}\,f(x)\leqslant f(x)$．凸函数 f 称为闭的，如果 $f=\mathrm{cl}\,f$．特别地，可以证明，正常凸函数 f 为闭的充要条件是：对每个实数 α，凸水平集 $\{x: x\in R^n,\ f(x)\leqslant\alpha\}$ 是闭的．

等价地说，正常凸函数的闭包运算直接地联系着集合的闭包运算．以记号 $\overline{C}$ 表示集合 C 的闭包，$\mathrm{cl}\,f$ 的上图象将是 f 的上图象的闭包，即(见[13])

$$P(\mathrm{cl}\,f)=\overline{P(f)}. \tag{4.65}$$

当然，上述关系也能用作正常凸函数闭包运算的定义．对于非正常凸函数，闭包运算具有十分简单的含义．如对每个 $x\in R^n$ 有 $f(x)=+\infty$，则 $\mathrm{cl}\,f=f$．然而，如对某些 $x\in R^n$ 有 $f(x)=-\infty$，

---

* 译注：这里的原文为"对给定点 $x^0\in R$ 成立 $h(x^0)\leqslant f(x^0)$"，这与(4.63)不符．

则 $L(f)=\emptyset$; 由于按约定，空集的上确界取为 $-\infty$，我们就对每个 $x\in R^n$ 得到 $\mathrm{cl}\,f(x)=-\infty$. 这样，闭的非正常凸函数只能是恒为 $+\infty$ 或 $-\infty$. 对正常凸函数，闭的概念与下半连续性的概念是等价的[13]. 类似地，正常凹函数是闭的，当且仅当它是上半连续的.

例 4.2.3

我们对于由(4.62)定义的凸函数进行闭包运算. 因为它是正常凸函数，我们只要对图 4.4* 中的上图象取"闭包". 这样，我们得到 f 的闭包为

$$\mathrm{cl}\,f=\begin{cases}(x)^2-1, & x\leqslant 1,\\ +\infty, & x>1,\end{cases} \tag{4.66}$$

并且

$$\mathrm{cl}\,f(1)=0. \tag{4.67}$$

换言之，此例中的闭包运算降低了 f 在它的有效区域边界上的值，直至该函数在 $\mathrm{ED}(f)$ 上变为连续. ▌

正如我们刚刚从一个简单的例子中所看到的，凸函数的闭包运算消去了在有效区域边界点的某种不连续性. 函数 $\mathrm{cl}\,f$ 与 f 在 $\mathrm{ED}(f)$ 内部的每一点上是一致的. 其原因是凸函数在它的有效区域的内部是连续的，这一点读者也许可以猜得. 在叙述稍更一般的结果之前，我们叙述一个非空凸集的相对内部的概念.

考察 R^n 中具有如下性质的子集 B：对任意两点 $x^1\in B$、$x^2\in B$ 和任何实数值 α，也有 $\alpha x^1+(1-\alpha)x^2\in B$. 具有这个性质的集合称为**仿射集**[13]. 单个点、直线及超平面是 R^n 中仿射集的例子. 设已给一个凸集 $C\subset R^n$，包含 C 的所有仿射集之交称为 C 的**仿射包**. 将 C 看作它的仿射包的子集而确定的 C 的内部，称为 C 的**相对内部**，以 $\mathrm{ri}\,C$ 表示.

例 4.2.4

设 $a<b$，考察点 $(a,0)$ 与 $(b,0)$ 之间的线段构成的凸集 $C\subset R^2$，即

$$C=\{(x_1,\ 0):a\leqslant x_1\leqslant b\}. \tag{4.68}$$

---

* 译注：原图有错。

如果我们将 C 看作 R^2 的子集，C 将没有内部，因为不能在 C 中找到 R^2 的开超球. 当然，C 的仿射包为整个 x_1 轴(一维)，C 的相对内部为

$$\mathrm{ri}\, C=\{(x_1,\ 0): a<x_1<b\}, \tag{4.69}$$

它正是把 C 看作 x_1 轴的子集时 C 的内部. ▌

注意，当凸集 $C\subset R^n$ 为 n 维时，就象是 R^2 中的正方形或 R^3 中的立方体时，C 的仿射包就是 R^n 本身，所以 C 的相对内部与 C 的内部一致.

以下两个定理的证明，需要另一些拓扑中的概念以及凸集和凸函数的进一步性质，这些并不直接与我们研究的主题有关. 因为这样的缘故，我们在此只叙述这些定理，有兴趣的读者可在[13]中找到证明.

下面的第一个定理涉及某种非正常凸函数. 这种凸函数或者到处取值 $+\infty$，或者在其有效区域的某些点取值 $-\infty$.

定理 4.16

设 f 是一个非正常凸函数，则对于它的有效区域的相对内部的一切 x，均有 $f(x)=-\infty$.

我们现在叙述一个主要结果，它推广了"开凸集上的实值凸函数必定连续"的结论，例如，可参见[2, 5].

定理 4.17

一个凸函数在其有效区域的相对内部中总是连续的。

由此定理，我们立即就有下列推论.

推论 4.18

R^n 上的实值凸函数必定处处连续.

在本节的结束部分，我们利用前面给出的关于凸集的分离定理，来证明含有凸函数的不等式组的解的存在性的某些结果. 下列结果属于 Fan, Glicksberg 和 Hoffman[6].

定理 4.19

设 $f_1, \cdots, f_m$ 为正常凸函数，非空凸集 $C\subset\bigcap_{i=1}^{m}\mathrm{ED}(f_i)$，则下列

不同情况中恰发生一种情况：

(i) 存在一个 $x^0\in C$，使

$$f_i(x^0)<0, \quad i=1, \cdots, m. \tag{4.70}$$

(ii) 存在不全为 0 的非负数 $\alpha_1, \cdots, \alpha_m$，使对一切 $x\in C$，均有

$$\sum_{i=1}^m \alpha_i f_i(x)\geqslant 0. \tag{4.71}$$

【证明】 假设(i)成立．对于至少有一个 $\alpha_i>0$ 的任何非负数 $\alpha_1, \cdots, \alpha_m$，可得

$$\sum_{i=1}^m \alpha_i f_i(x^0)<0. \tag{4.72}$$

因此(ii)不能成立．

现在假设(i)不成立．令

$$Y=\{y: y\in R^m, \exists x\in C \text{ 使 } f_i(x)<y_i, i=1, \cdots, m\}. \tag{4.73}$$

这样，每个 $y\in Y$ 至少有一个正分量．因为 f_i 是正常凸函数，集合 Y 是非空的．进一步还可以证明它是一个凸集．设 $y^1\in Y$、$y^2\in Y$ 及 $x^1\in C$、$x^2\in C$ 使

$$f_i(x^1)<y_i^1, \quad f_i(x^2)<y_i^2, \quad i=1, \cdots, m, \tag{4.74}$$

则对任何权向量 q，我们有 $(q_1x^1+q_2x^2)\in C$，并且，

$$f_i(q_1x^1+q_2x^2)\leqslant q_1f_i(x^1)+q_2f_i(x^2)<q_1y_i^1+q_2y_i^2, \quad i=1, \cdots, m. \tag{4.75}$$

这样，$(q_1y^1+q_2y^2)\in Y$．设 $N\subset R^m$ 为非正象限(它是凸集)，即

$$N=\{y: y\in R^m, y\leqslant 0\}. \tag{4.76}$$

集合 Y 和 N 是两个非空的互不相交的凸集，由定理 4.7，必定存在一个超平面来分离它们．这样，便存在一个非零向量 $\alpha\in R^m$ 及一个实数 b，使对任何 $y\in Y$，有

$$\sum_{i=1}^m \alpha_i y_i\geqslant b, \tag{4.77}$$

而对任何 $y\in N$，有

$$\sum_{i=1}^m \alpha_i y_i\leqslant b. \tag{4.78}$$

假如 $\alpha_k<0$, 取向量 $y=(0, \cdots, 0, y_k, 0, \cdots, 0)\in N$, 其中 y_k 是负的, 取值充分大, (4.78) 將被破坏. 这就是说, 对所有 $i=1, \cdots, m$, 均有 $\alpha_i\geqslant 0$, 并有 $b\geqslant 0$. 对每个 $x\in C$, 令 y 由下式定义:

$$y_i=f_i(x)+\varepsilon, \quad i=1, \cdots, m, \tag{4.79}$$

其中 ε 是一个任意的正数. 于是 $y\in Y$, 且

$$\sum_{i=1}^{m}\alpha_i f_i(x)+\sum_{i=1}^{m}\alpha_i\varepsilon\geqslant b\geqslant 0, \tag{4.80}$$

因此对一切 $x\in C$, 均有

$$\sum_{i=1}^{m}\alpha_i f_i(x)\geqslant 0, \tag{4.81}$$

故 (ii) 成立. ▌

Rockafellar[13] 用例子说明了: 条件 $C\subset\bigcap_i \mathrm{ED}(f_i)$ 或它的某种较弱的形式, 对于上述定理是必要的. 设

$$f_1(x)=\begin{cases}-(x)^{1/2}, & x\geqslant 0,\\ +\infty, & x<0,\end{cases}$$

$$f_2(x)=x,$$

而 $C=R$. 可知不存在 $x\in C$ 使 $f_1(x)<0$ 且 $f_2(x)<0$. 同时也不存在不全为 0 的非负数 α_1 和 α_2, 使 $\alpha_1 f_1(x)+\alpha_2 f_2(x)\geqslant 0$ 对一切 $x\in C$ 成立. 这样 (i) 与 (ii) 均不成立. 其原因是 $\mathrm{ED}(f_1)\cap\mathrm{ED}(f_2)$ 为非负象限, 而 C 不包含在其中.

对定理 4.19 可以作许多进一步的推广. 其中之一可看作 Farkas 引理[3] 的推广形式, 在本章的后面部分也将是有用的.

定理 4.20

假设 $f_0, f_1, \cdots, f_m$ 是正常凸函数, 并设非空凸集 $C\subset\bigcap_{i=0}^{m}\mathrm{ED}(f_i)$. 如果下列不等式组不具有解 $x\in C$:

$$f_0(x)<0, \tag{4.82}$$

$$f_i(x)\leqslant 0, \quad i=1, \cdots, m, \tag{4.83}$$

同时, 却存在一个 $x^0\in C$, 使得

$$f_i(x^0)<0, \quad i=1, \cdots, m. \tag{4.84}$$

那么，或者对一切 $x\in C$ 均有

$$f_0(x)\geqslant 0, \tag{4.85}$$

或者存在不全为 0 的非负数 $\alpha_1, \cdots, \alpha_m$, 使对一切 $x\in C$ 均有

$$f_0(x)+\sum_{i=1}^{m}\alpha_i f_i(x)\geqslant 0. \tag{4.86}$$

【证明】 如果(4.82)与(4.83)在 C 中无解，那么 (4.82) 与 $f_i(x)<0$, $i=1, \cdots, m$, 在 C 中也无解. 根据定理 4.19, 必存在不全为 0 的非负数 $\lambda_0, \lambda_1, \cdots, \lambda_m$, 使对一切 $x\in C$ 有

$$\lambda_0 f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)\geqslant 0. \tag{4.87}$$

不失一般性，我们可设 $\lambda_0+\lambda_1+\cdots+\lambda_m=1$. 如 $\lambda_0=1$, 则(4.85)成立. 如 $0<\lambda_0<1$, 将(4.87)式除以 λ_0, 并对 $i=1, \cdots, m$, 令 $\alpha_i=\lambda_i/\lambda_0$, 则(4.86)成立.

最后，如 $\lambda_0=0$, 则对一切 $x\in C$ 成立

$$\sum_{i=1}^{m}\lambda_i f_i(x)\geqslant 0, \tag{4.88}$$

由于 $x^0\in C$ 满足(4.84), 故(4.88)只能在 $\lambda_1=0, \cdots, \lambda_m=0$ 时成立. 这便与 λ_i 不全为 0 相矛盾. 也就表明，$\lambda_0=0$ 的情况不会发生. ▌

4.3 凸函数的微分性质

在本节的开头部分，我们回忆一下实函数微分中的一些概念，并用之于凸函数. 我们将任何向量 $y\in R^n$ 看成一个方向. 设实函数 f 定义在 R^n 的子集 S 上，点 x^0 在 S 的内部，则 f 在 x^0 沿方向 y 的导数，或者说，f 在 x^0 沿方向 y 的**方向导数**定义为

$$Df(x^0;\ y)=\lim_{t\to 0}\frac{f(x^0+ty)-f(x^0)}{t}, \tag{4.89}$$

这里假定上述极限存在.

这个定义的直接扩充如下：设 f 为定义在 R^n 上的任一函数，取值在扩充的实数集中. 设 x^0 是一个点，$f(x)$ 在 x_0 取有限值，又设 y 是一个方向. f 在 x^0 沿方向 y 的**右侧导数**定义为

$$D^+f(x^0;\ y)=\lim_{t\to 0^+}\frac{f(x^0+ty)-f(x^0)}{t}, \qquad (4.90)$$

这里假定上述极限存在，也包括极限为 $\pm\infty$，又 $t\to 0^+$ 表示 t 经过正数趋于 0．类似地，f 在 x^0 沿方向 y 的左侧导数定义为

$$D^-f(x^0;\ y)=\lim_{t\to 0^-}\frac{f(x^0+ty)-f(x^0)}{t}. \qquad (4.91)$$

对 $y=0$，$D^+f(x^0;\ 0)$ 与 $D^-f(x^0;\ x)$ 两者皆为 0．

读者容易验明，必成立

$$-D^+f(x^0;\ -y)=D^-f(x^0;\ y). \qquad (4.92)$$

如对某个 x^0 及 y 成立着

$$D^+f(x^0;\ y)=D^-f(x^0;\ y), \qquad (4.93)$$

则便有在(4.89)意义下的方向导数．设 $y=(1,\ 0,\ \cdots,\ 0)\in R^n$，则 f 在 x^0 沿 y 的方向导数正好是 $\partial f(x^0)/\partial x_1$，即 f 对于 x^1 的偏导数．类似地可得 f 关于 $x_2,\ \cdots,\ x_n$ 的各个偏导数．如果函数在点 x^0 可微[1]，则 f 在 x^0 沿一切方向 y 的方向导数是有限的，且由下式给出：

$$Df(x^0;\ y)=y^T\nabla f(x^0), \qquad (4.94)$$

如前，这里 $\nabla f(x^0)$ 为 f 的梯度在 x^0 所取的值．在叙述关于凸函数的方向导数的第一个结果之前，希望读者回忆一下正齐次的概念．设 f 是 R^n 上的函数，如果对每个 $x\in R^n$ 及正数 t 均有

$$f(tx)=tf(x), \qquad (4.95)$$

则称 $f(x)$ 为正齐次(1 次)的．

定理 4.21

设 f 是一个凸函数，$x\in R^n$，点 x 使 $f(x)$ 取值有限，则 f 在 x 沿任何方向 y 的右侧导数与左侧导数都存在．此外，D^+f 与 D^-f 是 y 的正齐次凸函数，且

$$D^+f(x;\ y)\geqslant D^-f(x;\ y). \qquad (4.96)$$

【证明】 设 $x^0\in R^n$ 是使 $f(x^0)$ 为有限的任何点，定义 $h(y)=f(x^0+y)-f(x^0)$．因为 $P(h)$ 可由凸集 $P(f)$ 平移而得，所以 h 为凸函数，且必有 $h(0)=0$．设 $t_2>t_1>0$，且 $q_1=t_1/t_2$、$q_2=$

$(t_2-t_1)/t_2$. 于是 $q_1>0$、$q_2>0$ 且 $q_1+q_2=1$, 所以得到

$$h(q_1t_2y+q_20)=h(t_1y)\leqslant q_1h(t_2y)+q_2h(0), \tag{4.97}$$

$$\frac{h(t_1y)}{t_1}\leqslant\frac{h(t_2y)}{t_2}. \tag{4.98}$$

上述不等式表明 $h(ty)/t$ 是 $t>0$ 的非减函数, 从而出现在(4.90)中的差商是 $t>0$ 的非减函数. 这样, 关于凸函数 f, 我们能对每个 y 写出下式:

$$D^+f(x^0;\ y)=\inf_{t>0}\frac{f(x^0+ty)-f(x^0)}{t}, \tag{4.99}$$

从而 $D^+f(x^0;y)$ 必存在(虽然它可能不是有限的). 因为 $D^+f(x^0;y)$ 对每个 y 存在, 由(4.92)即知 $D^-f(x^0;\ y)$ 也对每个 y 存在. 设 λ 是一个正数, 则

$$D^+f(x^0;\ \lambda y)=\lim_{t\lambda\to0^+}\frac{\lambda[f(x^0+t\lambda y)-f(x^0)]}{t\lambda} \tag{4.100}$$

$$=\lambda D^+f(x^0;\ y), \tag{4.101}$$

因此 D^+f 是正齐次的. 对 D^-f, 同一事实也正确. 由齐次性质可知 $D^+f(x^0;\ 0)=D^-f(x^0;\ 0)=0$.

以下我们证明 D^+f 是 y 的一个凸函数. 设 $(y^1,\ \alpha^1)\in P(D^+f)$、$(y^2,\ \alpha^2)\in P(D^+f)$, 于是

$$D^+f(x^0;\ y^1)\leqslant\alpha^1<+\infty, \tag{4.102}$$

$$D^+f(x^0;\ y^2)\leqslant\alpha^2<+\infty. \tag{4.103}$$

由 f 的凸性, 对任意 q 及 $t>0$, 我们便有

$$f(x^0+t(q_1y^1+q_2y^2))=f\left(\frac{1}{2}(x^0+2tq_1y^1)+\frac{1}{2}(x^0+2tq_2y^2)\right), \tag{4.104}$$

或

$$f(x^0+t(q_1y^1+q_2y^2))\leqslant\frac{1}{2}f(x^0+2tq_1y^1)+\frac{1}{2}f(x^0+2tq_2y^2). \tag{4.105}$$

注意, 由(4.102)与(4.103), f 在(4.105)右边的值对充分小的 t 不会是 $+\infty$; 因此不会出现不确定的表达式 $\infty+(-\infty)$. 这样, 从(4.105)两边减去 $f(x^0)$, 除以 t, 取当 $t\to0^+$ 时的极限, 我们得

到

$$D^+f(x^0;\ q_1y^1+q_2y^2)\leqslant D^+f(x^0;\ q_1y^1)+D^+f(x^0;\ q_2y^2). \tag{4.106}$$

将(4.102)和(4.103)分别乘以 q_1 和 q_2，利用 D^+f 的齐次性，并把所得式子相加，最后得

$$D^+f(x^0;\ q_1y^1+q_2y^2)\leqslant q_1\alpha^1+q_2\alpha^2. \tag{4.107}$$

这样，$(q_1y^1+q_2y^2,\ q_1\alpha^1+q_2\alpha^2)\in P(D^+f)$，这便是所需证明的。

由于 D^+f 的齐次性质及凸性，可知对每个 y，有

$$D^+f(x^0;\ y)+D^+f(x^0;\ -y)\geqslant D^+f(x^0;\ 0)=0. \tag{4.108}$$

再由(4.92)便得

$$D^+f(x^0;\ y)\geqslant D^-f(x^0;\ y). \tag{4.109}$$

▮

下一个要引入的概念是次梯度，在可微凸函数情形，它与通常的梯度有关，在更一般的情形，它与方向导数有关．向量 $\xi\in R^n$ 称为凸函数 f 在点 $x\in R^n$ 的次梯度，如果对一切 $y\in R^n$，均成立

$$f(y)\geqslant f(x)+\xi^T(y-x). \tag{4.110}$$

对于一个凸函数 f，在某点 x，可能发生三种情形：(a)没有向量使(4.110)成立；(b)有唯一的 ξ 满足(4.110)；(c)存在不止一个这样的向量 ξ．Rockafellar[18]已经透彻地研究了凸函数的微分性质，仿照他的工作，我们以 $\partial f(x)$ 表示凸函数 f 在 x 的所有次梯度的集合．次梯度的某些基本性质可以概括为：集合 $\partial f(x)$，亦称为 f 的次微分，是一个闭的凸集．$\partial f(x)$ 恰好就是单个向量 $\xi\in R^n$ 的充要条件为：凸函数 f 按通常的意义在 x 可微．并且，这时有 $\xi=\nabla f(x)$，也就是成立

$$\xi_j=\frac{\partial f(x)}{\partial x_j},\quad j=1,\ \cdots,\ n. \tag{4.111}$$

次梯度的特性可以用方向导数来表征，这在以下定理中可以看到．

定理 4.22

设在点 x 处 $f(x)$ 为有限，则向量 $\xi\in R^n$ 为凸函数 f 在点 x 的次梯度的充要条件为：对一切方向 z 均成立

$$D^+f(x;\ z)\geqslant\xi^Tz. \tag{4.112}$$

【证明】 设 ξ 是 f 在 x 的次梯度，于是它满足(4.110)．令 $y=x+tz$，这里 t 是一个正数，这样，对每个 $z\in R^n$ 及 $t>0$ 均有

$$f(x+tz)\geqslant f(x)+t\xi^Tz. \tag{4.113}$$

将(4.113)两边除以 t，并重新排列，我们对每个 $z\in R^n$ 及 $t>0$ 得到

$$\frac{f(x+tz)-f(x)}{t}\geqslant\xi^Tz. \tag{4.114}$$

注意 $D^+f(x;z)$ 是(4.114)中差商的下确界，便得不等式(4.112)．相反地，如果(4.112)对每个 $z\in R^n$ 成立，则利用相同的论证，可知(4.113)也成立．因此，(4.110)成立，这就表明 ξ 是 f 在 x 的次梯度．】

由上述两个定理，我们有以下推论．

推论 4.23

设 f 是 R^n 上的凸函数，并设 $f(x)$ 为有限．则对一切 $y\in R^n$ 均有

$$f(y)\geqslant f(x)+D^+f(x;\ y-x). \tag{4.115}$$

特别地，如果 f 在 x 可微，则

$$f(y)\geqslant f(x)+(y-x)^T\nabla f(x). \tag{4.116}$$

这个推论的证明留给读者．

推论 4.24

设 f 是 R^n 上的凸函数，并设 $f(x)$ 和 $f(y)$ 为有限．则成立

$$D^+f(y;\ y-x)\geqslant D^+f(x;\ y-x), \tag{4.117}$$

和

$$D^-f(y;\ y-x)\geqslant D^-f(x;\ y-x). \tag{4.118}$$

特别地，如果 f 在 x 和 y 为可微，则

$$(y-x)^T[\nabla f(y)-\nabla f(x)]\geqslant 0. \tag{4.119}$$

【证明】 由推论 4.23，我们有

$$f(y) \geqslant f(x) + D^+f(x; y-x), \tag{4.120}$$

$$f(x) \geqslant f(y) + D^+f(y; x-y), \tag{4.121}$$

这便得到

$$-D^+f(y; x-y) \geqslant D^+f(x; y-x). \tag{4.122}$$

并且由(4.92)与定理 4.21，有

$$D^+f(y; y-x) \geqslant D^-f(y; y-x) = -D^+f(y; x-y), \tag{4.123}$$

所以得到

$$D^+f(y; y-x) \geqslant D^+f(x; y-x). \tag{4.124}$$

关于左侧导数的类似结果的证明是相同的. ▌

定理 4.15 表明，函数 f 在 R^n 上的凸性等价于将它限制于 R^n 中任何直线段时的凸性. 因此，在某些情形，只需要研究在实直线 R 上的凸函数的性质. 从另一角度来看，这种研究也是很吸引人的，因为与多维情形比较起来，它的结果往往相当简明. 例如，在一维情形，f 定义在 R 上，f 在点 x 的所有右侧和左侧导数能够相应地由 $D^+f(x; 1)$ 与 $D^-f(x; 1)$ 算得，这是因为可以利用这些导数的齐次性质. 现在，我们要阐明这些导数是关于 x 的单调非减函数.

定理 4.25

设 f 是 R 上的一个凸函数，并设 x^1 和 x^2 是使 $f(x^1)$ 和 $f(x^2)$ 皆为有限的两点，$x^2 > x^1$，则成立

$$D^+f(x^2; 1) \geqslant D^-f(x^2; 1) \geqslant D^+f(x^1; 1) \geqslant D^-f(x^1; 1). \tag{4.125}$$

【证明】 上述第一个不等式和最后一个不等式从定理 4.21 即可得到. 中间一个不等式可以从以下关系式中得到：

$$D^-f(x^2; 1) \geqslant \frac{f(x^1) - f(x^2)}{x^1 - x^2} = \frac{f(x^2) - f(x^1)}{x^2 - x^1} \geqslant D^+f(x^1; 1). \tag{4.126}$$

▌

例 4.3.1

为了说明至此已得的概念和结果，考虑定义在 R 上的下列凸函数：

$$f(x)=\begin{cases}+\infty, & x<-1,\\ 2, & x=-1,\\ (x)^2, & -1<x\leqslant 0,\\ x, & 0\leqslant x\leqslant 1,\\ +\infty, & 1<x.\end{cases}\tag{4.127}$$

利用右侧导数和左侧导数的定义，并注意它们可以相应地由 $D^+f(x;\ 1)$ 与 $D\ f(x;\ -1)$ 算得，我们得到

$$D^+f(x;\ 1)=\begin{cases}\text{无意义}, & x<-1,\\ -\infty, & x=-1,\\ 2x, & -1<x<0,\\ 1, & 0\leqslant x<1,\\ +\infty, & x=1,\\ \text{无意义}, & 1<x.\end{cases}\tag{4.128}$$

和

$$D^-f(x;\ 1)=\begin{cases}\text{无意义}, & x<-1,\\ -\infty, & x=-1,\\ 2x, & -1<x\leqslant 0,\\ 1, & 0<x\leqslant 1,\\ \text{无意义}, & 1<x.\end{cases}\tag{4.129}$$

我们可以看到，对 $-1\leqslant x<0$ 及 $0<x<1$，有 $D^+f(x;1)=D^-f(x;1)$。而对 $x=0$ 和 $x=1$ 有 $D^+f(x;\ 1)>D^-f(x;\ 1)$。由定理 4.22，数 ξ 是 f 的次梯度的充要条件为：对一切 $z\in R$，成立

$$D^+f(x;\ z)\geqslant \xi z.\tag{4.130}$$

又因为单侧导数是正齐次的，我们得到

$$D^+f(x;\ z)=\begin{cases}zD^+f(x;\ 1), & z>0,\\ 0, & z=0,\\ -zD^+f(x;\ -1), & z<0.\end{cases}\tag{4.131}$$

根据上述两个关系式，我们可以得出，$\xi\in\partial f(x)$ 等价于

$$D^+f(x;1)\geqslant\xi\geqslant D^-f(x;1). \tag{4.132}$$

所以有

$$\partial f(x)=\begin{cases}\emptyset, & x\leqslant -1,\\ 2x, & -1<x<0,\\ \{\xi: 0\leqslant\xi\leqslant 1\}, & x=0,\\ 1, & 0<x<1,\\ \{\xi:\xi\geqslant 1\}, & x=1,\\ \emptyset, & 1<x.\end{cases} \tag{4.133}$$ ▌

在以下一些结果中，我们将限于注意实值的可微凸函数．这些结果中的大多数可以推广到按 § 4.2 开始部分所述的广泛意义下的凸函数，推广时要用到单侧导数（右侧和左侧）以及次梯度这些较一般的概念．然而，这种更为一般的结果的证明，往往变得相当复杂，这是由于需涉及到所考虑的函数在有效区域边界和外部的性质．由于这个原因，我们在本节的其余部分决定损失一些一般性来提出这些结果．然而，有兴趣的读者只要在难捉摸之处注意一下就能扩充这些结果．

定理 4.26

假设 f 是开区间 $D\subset R$ 上的实值可微函数，那么，当且仅当 f 为 D 上凸函数时，f 的一阶导数 f' 为 D 上的非减函数．

【证明】 若 f 为凸函数，由定理 4.25 即推得结论．现在，设 $x^1\in D$、$x^2\in D$，使得 $x^2>x^1$，且令 $x^3=q_1x^1+q_2x^2$，利用中值定理[1]，有

$$f(x^2)=f(x^3)+q_1(x^2-x^1)f'(\tilde{x}),\quad x^2\geqslant\tilde{x}\geqslant x^3, \tag{4.134}$$

$$f(x^3)=f(x^1)+q_2(x^2-x^1)f'(\tilde{\tilde{x}}),\quad x^3\geqslant\tilde{\tilde{x}}\geqslant x^1. \tag{4.135}$$

当 f' 是 R 上非减函数时，我们就有

$$f(x^3)\leqslant f(x^1)+q_2(x^2-x^1)f'(\tilde{x}). \tag{4.136}$$

将(4.136)乘以 q_1，将(4.134)乘以 $-q_2$，再加起来，得到

$$q_1f(x^3)-q_2f(x^2)\leqslant q_1f(x^1)-q_2f(x^3), \tag{4.137}$$

即

$$f(x^3)\leqslant q_1f(x^1)+q_2f(x^2), \tag{4.138}$$

这就表明，f 是一个凸函数． ∎

至此建立的大部分微分学的结果都是凸函数的必要条件．下一定理表明，在可微凸函数的特殊情形中，这些条件作为充分条件也很容易证明．

定理 4.27

设 f 是 R^n 上的实值可微函数，若对任何两点 $x^1\in R^n$, $x^2\in R^n$，有

$$f(x^2)\geqslant f(x^1)+(x^2-x^1)^T\nabla f(x^1), \tag{4.139}$$

则 f 在 R^n 上是凸的．

【证明】 设 x^1 和 x^2 是 R^n 中任意两点，且令 $x^3=q_1x^1+q_2x^2$，于是有

$$f(x^1)\geqslant f(x^3)+q_2(x^1-x^2)^T\nabla f(x^3), \tag{4.140}$$

$$f(x^2)\geqslant f(x^3)+q_1(x^2-x^1)^T\nabla f(x^3). \tag{4.141}$$

分别用 q_1 和 q_2 乘(4.141)和(4.140)，再加起来，我们得到

$$q_1f(x^1)+q_2f(x^2)\geqslant f(x^3), \tag{4.142}$$

从而 f 是凸的． ∎

推论 4.23 和定理 4.27 具有简明的几何意义．R^n 上的一个可微函数为凸函数的充要条件是：f 在点 x^0 的 Taylor 展开式的前二项，亦即如下线性函数：

$$f(x^0)+(x-x^0)^T\nabla f(x^0), \tag{4.143}$$

对任何 $x\in R^n$ 均小于等于 $f(x)$．

下一定理涉及凸函数的一阶偏导数的连续性．这个定理的证明需要其他背景材料，因此予以省略．有兴趣的读者可以参考 Fenchel[8]，Rockafellar[12] 或 Stoer、Witzgall[15]．

定理 4.28

设 f 是开凸集 $C\subset R^n$ 上的实值凸函数，若 f 在 C 上可微，则 f 在 C 上必有连续的一阶偏导数．

在结束本节时，我们引入关于二阶可微凸函数的一些结果．

定理 4.29

设 f 是开区间 $D\subset R$ 上的实值二阶可微函数，则 f 是 D 上凸函数的充要条件为：f 的二阶导数 f'' 在每个 $x\in D$ 上取值都是非负的.

【证明】 由定理 4.26，f 在 D 上为凸函数的充要条件是 f' 为非减函数，也就是，对每个 $x\in D$ 有 $f''(x)\geqslant 0$. ▐

现在，我们将上述结果推广到多维情形.

定理 4.30

设 f 是开凸集 $C\subset R^n$ 上的一个实值函数，它具有二阶连续偏导数. 则 f 是 C 上凸函数的充要条件为：在每个 $x\in D$ 算得的 f 的 Hesse 阵是半正定的. 这也就是说，在每个 $x\in D$，对一切 $y\in R^n$，我们有

$$y^T\nabla^2 f(x)y\geqslant 0. \tag{4.144}$$

【证明】 由定理 4.15，函数 f 是 C 上凸函数的充要条件为：对任何 $x^1\in C$、$x^2\in C$，函数

$$\phi(\lambda)=f(\lambda x^1+(1-\lambda)x^2) \tag{4.145}$$

对 $1\geqslant\lambda\geqslant 0$ 为凸函数. 由定理 4.29，上述条件等价于：函数

$$\phi''(\lambda)=\sum_{j=1}^{n}\sum_{k=1}^{n}(x_j^1-x_j^2)(x_k^1-x_k^2)\frac{\partial^2 f(\lambda x^1+(1-\lambda)x^2)}{\partial x_j\partial x_k} \tag{4.146}$$

对任何 $x^1\in C$、$x^2\in C$ 和 $0\leqslant\lambda\leqslant 1$ 均为非负. 令 $y=x^1-x^2$，因为 $x=\lambda x^1+(1-\lambda)x^2\in C$，我们可将(4.146)改写为：对任何 $y\in R^n$ 及 $x\in C$，成立

$$\sum_{j=1}^{n}\sum_{k=1}^{n}y_j y_k\frac{\partial^2 f(x)}{\partial x_j\partial x_k}\geqslant 0. \tag{4.147}$$

▐

这里列入一点注意事项. 将定理 4.30 中“半正定”一词改为“正定”，这个定理不能整个地加强到严格凸函数的情形. 事实上，读者容易找到 Hesse 阵不是正定的严格凸函数的例子.

然而，相反情形的状况要好些，这就是说，在定理的假设之下，

正定 Hesse 阵必定蕴涵函数的严格凸性．关于这些论点的透彻研究，读者可参见[4]．

4.4 凸函数的极值

早已指出，凸函数及其推广在非线性规划中起着主要作用．凸函数的根本重要性在于下面概述的某些基本性质．

定理 4.31

设 f 是 R^n 上的正常凸函数，则 f 的任何局部极小值点必是 f 在 R^n 上的一个整体极小值点．

【证明】 如果 x^* 是一个局部极小值点，那么对于充分小的邻域 $N_\delta(x^*)$ 中的一切 x，均有

$$f(x)\geqslant f(x^*). \tag{4.148}$$

令 z 是 R^n 中的任一点，对充分小的 λ，$0<\lambda<1$，就有$((1-\lambda)x^*+\lambda z)\in N_\delta(x^*)$，从而

$$f((1-\lambda)x^*+\lambda z)\geqslant f(x^*). \tag{4.149}$$

由于 f 是正常凸函数，有

$$(1-\lambda)f(x^*)+\lambda f(z)\geqslant f((1-\lambda)x^*+\lambda z). \tag{4.150}$$

将上述两个不等式相加，移项后除以 λ，我们得到

$$f(z)\geqslant f(x^*), \tag{4.151}$$

这就是说，x^* 是一个整体极小值点．】

假设在§3.1中定义的问题(P)中，函数 $g_1, \cdots, g_m$ 皆为凹函数，又 $h_1, \cdots, h_p$ 皆为线性函数，于是，能行集 X 是一个凸集．如果要极小化的目标函数是一个正常凸函数，我们可以定义一个新的目标函数如下：

$$\hat{f}(x)=\begin{cases} f(x), & x\in X, \\ +\infty, & x\bar{\in} X, \end{cases} \tag{4.152}$$

并且，$\hat{f}(x)$也将是 R^n 上的正常凸函数，它与 f 在 X 上相同．根据前述定理我们得出，$\hat{f}$ 的每个局部极小值点也是一个整体极小值点，或者说，如 X 非空，f 在某点 $x\in X$ 达到的任何局部极小也是 f 在整个 X 的整体极小．

定理 4.32

设 f 是 R^n 上的正常凸函数,并设 $X\subset R^n$ 是一个凸集,则 f 在 $x\in X$ 达到的任何局部极小必是 f 在整个 X 上的整体极小.

注意,一般地,凸函数的极小值可以在多于一点处达到. 我们现在要证明,一个正常凸函数的极小点的集合必是一个凸集. 首先,我们有下列结果.

引理 4.33

设 f 是 R^n 上的凸函数, α 是一个实数, f 的水平集定义为

$$S(f, \alpha)=\{x: x\in R^n, f(x)\leqslant\alpha\}, \tag{4.153}$$

则对任何 α, $S(f, \alpha)$ 均为凸集.

【证明】 假设 $x^1\in S(f, \alpha)$ 和 $x^2\in S(f, \alpha)$. 于是便有

$$f(q_1x^1+q_2x^2)\leqslant q_1f(x^1)+q_2f(x^2) \tag{4.154}$$

$$\leqslant q_1\alpha+q_2\alpha=\alpha. \tag{4.155}$$

所以 $S(f, \alpha)$ 是凸集. ▋

定理 4.34

设 f 是 R^n 上的凸函数,则使 f 达到极小值的点的集合必是凸集.

【证明】 设 α^* 是 f 在极小值点的值. 于是,集合 $\{x: x\in R^n, f(x)\leqslant\alpha^*\}$ 恰是使 f 达到极小值的点的集合,由上述引理它必是一个凸集. ▋

另外一个结果在某些应用中也十分重要,在这里提出是值得的.

推论 4.35

设 f 是严格凸函数,它定义在凸集 $X\subset R^n$ 上. 如果 f 在 X 上能达到它的极小值,则它必在 X 的唯一点达到.

【证明】 假设极小值在两个不同的点达到,这两个点是 $x^1\in X$ 与 $x^2\in X$,且令 $f(x^1)=f(x^2)=\alpha$. 由定理 4.34 可知,对任何 q 我们有 $f(q_1x^1+q_2x^2)=\alpha$,这便与 f 为严格凸函数矛盾. ▋

在许多应用中,当要寻求可微函数的极小值点时,人们就去找出函数的逗留点——使梯度为 0 的点. 在凸函数的情形,这种处

理能被证明是完全正当的.

定理 4.36

假设 f 是一个凸函数，那么，当且仅当 f 在 x^* 达到极小值时，有 $0\in\partial f(x^*)$.

【证明】 由次梯度的定义，$0\in\partial f(x^*)$ 等价于对一切 $y\in R^n$ 成立

$$f(y)\geqslant f(x^*), \tag{4.156}$$

也就是说，x^* 是 f 的一个极小值点. 】

推论 4.37

设 f 是 R^n 上的可微凸函数，那么，当且仅当 f 在 x^* 达到极小值时成立

$$\nabla f(x^*)=0. \tag{4.157}$$

【证明】 当且仅当 f 在 x 可微时，$\partial f(x)=\{\nabla f(x)\}$. 所以这个推论是上述定理的直接结果. 】

如果我们将 R^n 代之以某个开凸集 $X\subset R^n$，而使 $x^*\in X$，则上述推论一般仍保持正确. 它也表明，在寻找凸函数的无约束极小时，在梯度为 0 的点不需要再检验二阶条件. 当然，这个结论也能从定理 4.30 推得.

4.5 凸规划的最优性条件

这里我们考察凸规划，它是第 3 章中引入的一般非线性规划问题(P)的特殊情形. 那里所导出的最优性条件对于凸规划来说将变得较为简单. 下面，就考察下列称为凸规划的非线性规划：

(CP)
$$\min f(x) \tag{4.158}$$

受限制于约束

$$g_i(x)\geqslant 0,\quad i=1,\ \cdots,\ m, \tag{4.159}$$

$$h_j(x)=0,\quad j=1,\ \cdots,\ p, \tag{4.160}$$

其中 f 是 R^n 上的正常凸函数，g_i 是正常凹函数，h_j 是下列形式的线性(仿射)函数：

$$h_j(x)=\sum_{k=1}^{n}a_{jk}x_k-b_j. \tag{4.161}$$

这样的规划所以被称为凸规划，是因为目标函数是凸函数，并且满足(4.159)中不等式和(4.160)中等式的所有 $x\in R^n$ 的集合是一个凸集(见练习 4.Q). 注意，一般地，当 h 是非线性的凸函数或凹函数时，满足方程 $h(x)=0$ 的点 $x\in R^n$ 的集合不是一个凸集.

我们现在要表明，加以可微性的适当假设，定理 3.8 中叙述的关于最优性的 Kuhn-Tucker 必要条件，用于凸规划时也是充分条件.

定理 4.38

假设函数 f 和 $g_1, \cdots, g_m$ 分别为 R^n 上实值可微的凸函数和凹函数，并设 $h_1, \cdots, h_p$ 是线性函数. 若存在向量 x^*、λ^*、μ^*，其中 x^* 满足(4.159)和(4.160)，且成立

$$\nabla f(x^*)-\sum_{i=1}^{m}\lambda_i^*\nabla g_i(x^*)-\sum_{j=1}^{p}\mu_j^*\nabla h_j(x^*)=0, \tag{4.162}$$

$$\lambda_i^*g_i(x^*)=0, \quad i=1, \cdots, m, \tag{4.163}$$

$$\lambda^*\geqslant 0, \tag{4.164}$$

则 x^* 是(CP)的整体最优点.

【证明】 设 x 是满足(4.159)和(4.160)的任一点. 于是

$$f(x)\geqslant f(x)-\sum_{i=1}^{m}\lambda_i^*g_i(x)-\sum_{j=1}^{p}\mu_j^*h_j(x). \tag{4.165}$$

将推论 4.23 用于 f、g_i 和 h_j，并利用(4.165)，我们得到

$$\begin{aligned}f(x)\geqslant f(x^*)&+(x-x^*)^T\nabla f(x^*)-\sum_{i=1}^{m}\lambda_i^*g_i(x^*)\\&-\sum_{i=1}^{m}\lambda_i^*(x-x^*)^T\nabla g_i(x^*)-\sum_{j=1}^{p}\mu_j^*h_j(x^*)\\&-\sum_{j=1}^{p}\mu_j^*(x-x^*)^T\nabla h_j(x^*).\end{aligned} \tag{4.166}$$

重新排列上式，就得到

$$\begin{aligned}f(x)\geqslant f(x^*)&-\sum_{i=1}^{m}\lambda_i^*g_i(x^*)-\sum_{j=1}^{p}\mu_j^*h_j(x^*)\\&+(x-x^*)^T\left[\nabla f(x^*)-\sum_{i=1}^{m}\lambda_i^*\nabla g_i(x^*)-\sum_{j=1}^{p}\mu_j^*\nabla h_j(x^*)\right],\end{aligned} \tag{4.167}$$

由(4.160)、(4.162)和(4.163)，便可得到

$$f(x) \geqslant f(x^*). \tag{4.168}$$

∎

由这个结果显然可知，在同样的假设下，定理3.4中叙述的Fritz John条件在$\lambda_0^*>0$时也是充分条件.

在一般情形的最优性必要条件中所假设的正则性条件(3.71)，在凸规划中可以用更为简单、更易计算、显然是更强的条件来代替，该条件属于Slater[12, 14]. 在第3章中所定义的一般的数学规划(P)称为是强相容的，如果存在一点$x^0 \in R^n$满足

$$g_i(x^0) > 0, \quad i=1, \cdots, m, \tag{4.169}$$

$$h_j(x^0) = 0, \quad j=1, \cdots, p. \tag{4.170}$$

在这些条件下，也说这个规划满足Slater条件. 我们给予(CP)约束的补充条件是：线性函数$h_j(x)=(a^j)^T x - b_j$的系数向量$a^j=(a_{j1}, \cdots, a_{jn})^T$为线性无关. 这并不是一个实质性的限制，因为，如果系数向量之间线性相关，我们总能利用对这些向量的初等代数运算，在不改变问题的能行集的条件下，使一个或若干个方程变成一切$x \in R^n$均满足的方程$0^T x = 0$，其中左边的0是分量皆为0的n维向量. 照此办理，就可以消去所有线性相关的约束.

现在，我们能用相当直接的方法，证明凸规划(CP)的最优性的Kuhn-Tucker必要条件.

定理4.39

假设函数f和$g_1, \cdots, g_m$分别为R^n上实值可微的凸函数和凹函数. 设

$$h_j(x) = \sum_{k=1}^{n} a_{jk} x_k - b_j, \quad j=1, \cdots, p \tag{4.171}$$

而向量$a^j=(a_{j1}, \cdots, a_{jn})$，$j=1, \cdots, p$为线性无关. 又假设问题(CP)为强相容的. 如果x^*是(CP)的解，那么必存在向量$\lambda^*=(\lambda_1^*, \cdots, \lambda_m^*)$和$\mu^*=(\mu_1^*, \cdots, \mu_p^*)$，使下列各式成立：

$$\nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^{p} \mu_j^* \nabla h_j(x^*) = 0, \tag{4.172}$$

$$\lambda_i^* g_i(x^*)=0, \quad i=1, \cdots, m, \tag{4.173}$$

$$\lambda^* \geqslant 0. \tag{4.174}$$

【证明】 在关于 f、g_i 和 h_j 的假设下，根据定理 4.28 可知定理 3.4 的假设亦成立。所以必存在向量 $\alpha^*=(\alpha_0^*, \alpha_1^*, \cdots, \alpha_m^*)$ 和 $\beta^*=(\beta_1^*, \cdots, \beta_p^*)$ 使

$$\alpha_0^* \nabla f(x^*)-\sum_{i=1}^{m} \alpha_i^* \nabla g_i(x^*)-\sum_{j=1}^{p} \beta_j^* \nabla h_j(x^*)=0, \tag{4.175}$$

$$\alpha_i^* g_i(x^*)=0, \quad i=1, \cdots, m, \tag{4.176}$$

$$(\alpha^*, \beta^*) \neq 0, \quad \alpha^* \geqslant 0. \tag{4.177}$$

如果 $\alpha_0^*>0$，令 $\lambda_i^*=\alpha_i^*/\alpha_0^*$，$\mu_j^*=\beta_j^*/\alpha_0^*$，这定理就得到证明。我们现在要证明，$\alpha_0^*=0$ 的假定必导致矛盾。

设 $\alpha_0^*=0$ 和 $(\alpha_1^*, \cdots, \alpha_m^*) \neq 0$，又设 x^0 满足(4.169)和(4.170)。于是，由于 g_i 为凹函数，有

$$\sum_{i=1}^{m} \alpha_i^* g_i(x^0) \leqslant \sum_{i=1}^{m} \alpha_i^* g_i(x^*)+\sum_{k=1}^{n}(x_k^0-x_k^*)\left[\sum_{i=1}^{m} \alpha_i^* \frac{\partial g_i(x^*)}{\partial x_k}\right], \tag{4.178}$$

且由(4.175)和(4.176)，就有

$$\sum_{i=1}^{m} \alpha_i^* g_i(x^0) \leqslant \sum_{k=1}^{n}(x_k^0-x_k^*)\left[-\sum_{j=1}^{p} \beta_j^* \frac{\partial h_j(x^*)}{\partial x_k}\right]. \tag{4.179}$$

由于(线性的)h_j 是凹函数，因此有

$$\sum_{i=1}^{m} \alpha_i^* g_i(x^0) \leqslant \sum_{j=1}^{p} \beta_j^*[h_j(x^*)-h_j(x^0)]=0, \tag{4.180}$$

便与(4.169)相矛盾，这是因为，若对于 $i=1, \cdots, m$，有 $g_i(x^0)>0$，且 $(\alpha_1^*, \cdots, \alpha_m^*) \neq 0$，则必有

$$\sum_{i=1}^{m} \alpha_i^* g_i(x^0)>0. \tag{4.181}$$

现在设 $\alpha^*=0$，便必定有 $\beta^* \neq 0$。于是由(4.175)，我们有

$$\sum_{j=1}^{p} \beta_j^* \nabla h_j(x^*)=0, \tag{4.182}$$

或

$$\sum_{j=1}^{p} \beta_j^* a_{jk}=0, \quad k=1, \cdots, n, \tag{4.183}$$

这意味着向量 a^j 线性相关，又得到矛盾。 】

在要结束本章时，我们就凸规划情形给出定理 3.11 的逆定理．我们将证明，凸规划的每个解必是它相应的 Lagrange 式的鞍点．然而，还要先对定理 4.20 作出一个稍微不同的变形．

定理 4.40

设 $g_1, \cdots, g_m$ 是正常凹函数，又 $h_1, \cdots, h_p$ 是线性函数．设 $C\subset R^n$ 是非空凸集，使

$$C\subset\bigcap_{i=1}^{m}\mathrm{ED}(g_i). \tag{4.184}$$

如果不存在满足下列条件的 $x\in C$：

$$g_i(x)>0,\quad i=1,\ \cdots,\ m, \tag{4.185}$$

$$h_j(x)=0,\quad j=1,\ \cdots,\ p, \tag{4.186}$$

那么必存在不全为 0 的数 $\alpha_1, \cdots, \alpha_m, \beta_1, \cdots, \beta_p$，对 $i=1, \cdots, m$ 有 $\alpha_i\geqslant 0$，且对任何 $x\in C$，有

$$\sum_{i=1}^{m}\alpha_i g_i(x)+\sum_{j=1}^{p}\beta_j h_j(x)\leqslant 0. \tag{4.187}$$

此定理的证明类似于定理 4.20 的证明，留给读者．现在我们有下列推广的 Kuhn-Tucker 鞍点定理．

定理 4.41

设函数 $f, g_1, \cdots, g_m, h_1, \cdots, h_p$ 满足定理 4.39 的假设，也可以不满足其中的可微性假设，又设(CP)是强相容的．如果 x^* 是(CP)的解，那么必存在向量 $\lambda^*=(\lambda_1^*, \cdots, \lambda_m^*)$ 和 $\mu^*=(\mu_1^*, \cdots, \mu_p^*)$，使 (x^*, λ^*, μ^*) 是 §3.3 中定义的问题(S)的解，亦即，$\lambda^*\geqslant 0$ 且

$$L(x^*, \lambda, \mu)\leqslant L(x^*, \lambda^*, \mu^*)\leqslant L(x, \lambda^*, \mu^*). \tag{4.188}$$

同时，还满足

$$\lambda_i^* g_i(x^*)=0,\quad i=1,\ \cdots,\ m. \tag{4.189}$$

【证明】 下列系统不具有解 $x\in R^n$：

$$f(x^*)-f(x)>0, \tag{4.190}$$

$$\begin{aligned} g_i(x)\geqslant 0,&\quad i=1,\ \cdots,\ m,\\ h_j(x)=0,&\quad j=1,\ \cdots,\ p. \end{aligned} \tag{4.191}$$

根据定理 4.20，就必定存在不全为 0 的数 $\alpha_0\geqslant 0, \alpha_1\geqslant 0, \cdots,$

$\alpha_m \geqslant 0$, β_1, $\cdots$, β_p, 使对任何 x, 成立：

$$\alpha_0(f(x^*)-f(x))+\sum_{i=1}^{m}\alpha_i g_i(x)+\sum_{j=1}^{p}\beta_j h_j(x)\leqslant 0, \quad (4.192)$$

即

$$\alpha_0 f(x^*)\leqslant \alpha_0 f(x)-\sum_{i=1}^{m}\alpha_i g_i(x)-\sum_{j=1}^{p}\beta_j h_j(x). \quad (4.193)$$

设 $\alpha_0>0$. 于是，将 (4.193) 除以 α_0, 且对 $i=1, \cdots, m$ 令 $\lambda_i^*=\alpha_i/\alpha_0$, 对 $j=1, \cdots, p$ 令 $\mu_j^*=\beta_j/\alpha_0$, 则对任何 $x\in R^n$ 成立：

$$f(x^*)\leqslant L(x, \lambda^*, \mu^*). \quad (4.194)$$

由于 $g_i(x^*)\geqslant 0$ 和 $h_j(x^*)=0$, 于是对任何 $\lambda\in R^m$, $\lambda\geqslant 0$ 及 $\mu\in R^p$, 我们有

$$f(x^*)-\sum_{i=1}^{m}\lambda_i g_i(x^*)-\sum_{j=1}^{p}\mu_j h_j(x^*)=L(x^*, \lambda, \mu)\leqslant f(x^*). \quad (4.195)$$

将 x^*, λ^*, μ^* 代入(4.194) 和 (4.195)中各处，我们得到 $f(x^*)=L(x^*, \lambda^*, \mu^*)$. 由这个结果转过来便可推出，对 $i=1, \cdots, m$ 成立 $\lambda_i^* g_i(x^*)=0$. 因此，在 $\alpha_0>0$ 的情形，定理的证明便完成了.

现在设 $\alpha_0=0$. 于是对任何 $x\in R^n$, 有

$$\sum_{i=1}^{m}\alpha_i g_i(x)+\sum_{j=1}^{p}\beta_j h_j(x)\leqslant 0. \quad (4.196)$$

因为(CP)是强相容的，我们必有 $\alpha_1=\alpha_2=\cdots=\alpha_m=0$. 于是，$\beta\neq 0$, 且对任何 $x\in R^n$ 成立

$$\sum_{k=1}^{n}\left(\sum_{j=1}^{p}\beta_j a_{jk}\right)x_k\leqslant\sum_{j=1}^{p}\beta_j b_j. \quad (4.197)$$

如果对某个 k_0, $1\leqslant k_0\leqslant p$, 有 $\sum_{j=1}^{p}\beta_j a_{jk_0}\neq 0$, 我们可以选取 x_{k_0} 使 $|x_{k_0}|$ 充分大，又 $k\neq k_0$ 时选取 $x_k=0$, 这便破坏了(4.197). 所以

$$\sum_{j=1}^{p}\beta_j a_{jk}=0, \quad k=1, \cdots, n. \quad (4.198)$$

因为 $\beta\neq 0$, 上述方程便与 $a^j=(a_{j1}, \cdots, a_{jn})$ 为线性无关的假设相矛盾. 所以 α_0 不能为 0. ∎

练　习

4.A. 用一个例子说明凸集的和集不一定为凸集. 给出一些凸集的例子，它们的余集亦为凸集.

4.B. 证明定理 4.3.

4.C. 设 $C\subset R^p$ 是一个凸集，证明下列集合 $\overline{X}$ 为 R^n 中的凸集:

$$\overline{X}=\{x: x\in R^n,\ x=A\rho,\ \rho\in C\}, \tag{4.199}$$

其中 A 是给定的 $n\times p$ 实矩阵.

4.D. 给定 R^n 上的线性(仿射)函数

$$h(x)=a^Tx+b, \tag{4.200}$$

证明它既是凸函数，又是凹函数.

4.E. 设 f 是凸集 $C\subset R^n$ 上连续的实值函数，对任何 $x^1\in C$ 和 $x^2\in C$，均有

$$f\left(\frac{1}{2}x^1+\frac{1}{2}x^2\right)\leqslant\frac{1}{2}f(x^1)+\frac{1}{2}f(x^2), \tag{4.201}$$

证明 f 是凸函数.

4.F. 许多有名的不等式能够通过凸性来证明. 利用定理 4.9 导出著名的算术-几何平均值不等式:

$$\sum_{i=1}^{m}\alpha_i a_i\geqslant\prod_{i=1}^{n}(a_i)^{\alpha_i}, \tag{4.202}$$

其中 a_i 是已给正数，α_i 是任意的非负数，满足

$$\sum_{i=1}^{m}\alpha_i=1. \tag{4.203}$$

4.G. 证明定理 4.10.

4.H. 完成定理 4.15 的证明.

4.I. 设 f 是 R^n 上实值的正齐次函数，证明 f 为凸函数的充要条件是：对任何 $x\in R^n$ 和 $y\in R^n$，成立

$$f(x+y)\leqslant f(x)+f(y). \tag{4.204}$$

4.J. 设 f 是 R^n 上的实值凸函数. 证明：如果 f 在某个 $x\in R^n$ 处具有整体极大值，则对一切 $x\in R^n$，$f(x)$ 为常数.

4.K. 点 x^0 属于凸集 $C\subset R^n$，如果不存在点 $x^1\in C$ 和 $x^2\in C$，$x^1\neq x^2$，使 x^0 可以写成 $x^0=\lambda x^1+(1-\lambda)x^2$，$0<\lambda<1$，则称 x^0 为 C 的一个顶点. 证明：对于紧致凸集 $K\subset R^n$，R^n 上的实值凸函数在 K 上的极大值必在 K 的某个顶点达到. 可以利用下列结果：R^n 中任何紧致凸集是它的顶点的凸包[15].

4.L. (a) 证明 R^n 上正常凸函数 f 为闭的充要条件是：对任何实数 α，集合 $\{x:x\in R^n,\ f(x)\leqslant\alpha\}$ 是闭集.

(b) 证明：当且仅当 f 为正常凸函数时，$\mathrm{cl}f$ 为正常凸函数.

4.M. 设 f 是凸函数，在 $x^0\in R^n$ 取值为有限，且在 x^0 存在右侧和左侧导数. 证明对任何 y，有

$$-D^+f(x^0;\ -y)=D^-f(x^0;\ y). \tag{4.205}$$

4.N. 证明：若 f 是凸函数，则 $\partial f(x)$ 是闭凸集. 并证明：若 $\partial f(x^0)$ 为非空，则 $\mathrm{cl}f(x^0)=f(x^0)$. 设 $f(x)=\|x\|$，证明 f 为凸函数，并对任何 $x\in R^n$，计算 $\partial f(x)$.

4.O. 证明推论 4.23.

4.P. 设实值函数 f 为

$$f(x)=\sum_{j=1}^{n}x_j\ln x_j-\sum_{j=1}^{n}x_j\ln\left(\sum_{j=1}^{n}x_j\right), \tag{4.206}$$

证明 f 在下列集合上为凸函数：

$$C(\alpha)=\{x:x\in R^n,\ x_1+\cdots+x_n=\alpha,\ x>0,\ \alpha>0\}. \tag{4.207}$$

又问：f 还是严格凸函数吗？

4.Q. 证明：(CP)的能行集，即满足(4.159)和(4.160)的 $x\in R^n$ 的集合必是闭凸集.

4.R. 考察下列非线性规划：

$$\text{(NLP)}\qquad \min f(x) \tag{4.208}$$

受限制于

$$Ax=b, \tag{4.209}$$

$$x\geqslant 0, \tag{4.210}$$

其中，f 是 R^n 上的实值连续可微的凸函数，A 是已给的 $m\times n$ 实矩阵，$m<n$，b 是已给的 n 维向量.

求解上述规划的途径之一是，寻找一个 x^*，既满足前面的约束，又使下列线性规划的一个最优解正好为 $y=x^*$：

$$\text{(LP)}\qquad \min \bar{f}(y)=y^T\nabla f(x^*) \tag{4.211}$$

受限制于

$$Ay=b, \tag{4.212}$$

$$y\geqslant 0. \tag{4.213}$$

试证明，如果 $y=x^*$ 是(LP)的解，则 x^* 是(NLP)的解.

参考文献

1. BARTLE, R. G., *The Elements of Real Analysis*, 2nd ed., John Wiley & Sons, New York, 1976.
2. BERGE, C., *Topological Spaces*, Oliver & Boyd Ltd., Edinburgh, 1963.
3. BERGE, C., and A. GHOUILA-HOURI, *Programming, Games and Transportation Networks*, Methuen & Co. Ltd., London, 1965.
4. BERNSTEIN, B., and R. A. TOUPIN, "Some Properties of the Hessian Matrix of a Strictly Convex Function," *J. für die Reine und Angew. Math.*, **210**, 67–72 (1962).
5. EGGLESTON, H. G., *Convexity*, Cambridge University Press, Cambridge, England, 1958.
6. FAN, K., I. GLICKSBERG, and A. J. HOFFMAN, "Systems of Inequalities Involving Convex Functions," *Proc. Am. Math. Soc.*, **8**, 617–622 (1957).
7. FARKAS, J., "Über die Theorie der Einfachen Ungleichungen," *J. für die Reine und Angew. Math.*, **124**, 1–27 (1902).
8. FENCHEL, W., *Convex Cones, Sets and Functions*, Mimeographed lecture notes, Princeton University, Princeton, N.J. 1951.
9. HÖLDER, O., "Über einen Mittelwertsatz," *Göttinger Nachrichten*, pp. 38–47 (1889).
10. JENSEN, J. L. W. V., "Sur les Fonctions Convexes et les Inégalités Entre les Valeurs Moyennes," *Acta Math.*, **30**, 175–193 (1906).
11. KUHN, H. W., and A. W. TUCKER, "Nonlinear Programming," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman (Ed.), University of California Press, Berkeley, Calif., 1951.
12. MANGASARIAN, O. L., *Nonlinear Programming*, McGraw-Hill Book Co., New York, 1969.
13. ROCKAFELLAR, R. T., *Convex Analysis*, Princeton University Press, Princeton, N.J., 1970.
14. SLATER, M., "Lagrange Multipliers Revisited: A Contribution to Nonlinear Programming," Cowles Commission Discussion Paper, Math. 403, November 1950.
15. STOER, J., and C. WITZGALL, *Convexity and Optimization in Finite Dimensions I*, Springer-Verlag, Berlin, 1970.
16. VALENTINE, F. A., *Convex Sets*, McGraw-Hill Book Co., New York, 1964.

第 5 章

非线性凸规划的对偶性

在讨论非线性规划问题的对偶性时，必须从线性规划说起，在那里得到了数学规划中第一个和最完整的对偶性结果．按 Dantzig[5]，对偶性的概念首先由 Von Neumann 在 1947 年引进线性规划，后来由 Gale、Kuhn、Tucker[16] 表达成精确的形式．对偶性的思想是，对每个线性规划，称为原有规划，联系着另一个线性规划，称为对偶规划．对偶线性规划有一些有趣的性质，这些性质不仅从理论的观点看是精致的，而且对计算方面的目的和经济学的解释来说也是有意义的．关于线性规划中对偶性的综合研究，读者可参考 Dantzig[5] 和 Kuhn、Tucker[38]．

原有-对偶线性规划最重要的性质如下：

1．如果原有规划是求一个线性函数在一组线性约束下的极小，那么对偶规划是求另一个线性函数在另一组线性约束下的极大．给了一个(原有)线性规划，它的对偶是容易仅由原有规划的参数来表示的．在对偶规划中不出现原有变量，反之亦然．原有规划与对偶规划之间的关系是“对合”的，即对偶的对偶又是原有规划．

2．对一切原有规划的能行解和对偶规划的能行解，原有目标函数（要求其极小值）的值总是大于或等于对偶目标函数（要求其极大值）的值．如原有规划有最优解，则其对偶也有最优解，且这两个目标函数在各自的最优点处有同样的值．知道了原有规划或对偶规划的最优解，就能够得到另一个规划的最优解而无须对它另外求解．所以，如果对偶规划在计算上更吸引人的话，就可以通过求解对偶规划来求解原有规划．

3．线性规划在经济理论中起重要作用．一个最优对偶解可看作是原有规划的 Lagrange 乘子向量或隐蔽价格，反之亦然．很

多经济问题可以叙述成线性规划，这时对偶规划就转而代表相关的经济问题。线性规划和经济学的关系的讨论可在 [9] 中找到.

线性规划中的优美的对偶性定理，以及 Kuhn 和 Tucker 的里程碑式的文章[32] 鼓励很多学者把对偶性的概念推广到非线性规划问题. 但这方面早期结果是不令人鼓舞的. 虽然已经表明，可以对某些非线性规划表述一个相应的对偶规划，且有少量相似于线性规划的结果，但很快就明白，为了得到有意义的结果，只能考虑凸的原有规划. 不幸，如早期著作[10, 21, 24, 36]所表述的那样，这种规划的对偶通常是非凸规划，且在原有规划与对偶规划之间不存在线性情形下的“对合”，在那里，对偶的对偶是原有规划.

六十年代后期，凸规划的更完全的对偶理论出现了，它的一般特征相似于上面列举的原有-对偶线性规划的特征. 这理论可循不同线索导出. 这里所选的一种方式是基于共轭函数概念的，后者由 Fenchel[13]引进，又由 Rockafellar 在一系列著作中发展完善的，这些著作中的一些列在本章末尾的参考文献中. Rockafellar 关于凸对偶性的大部分结果概括在[32]中. 这个理论是很新的且具有极大潜力，但它在非线性最优化上的实际影响还未被充分理解. 本章给出凸规划的对偶性理论的要点.

按照现代的方式，非线性凸规划将结合问题参数的某种扰动来表示. 这种扰动通常代表规划中参数的微小改变，从理论和实际的观点来看都是重要的. 例如，十分重要的是要知道，对一个最优解存在、目标函数的极小值为有限的给定的规划问题，参数的微小改变可以引起目标函数值无限陡地减小. 所以，除了关于能行解和最优解的存在性及其特征这样的标准问题之外，这种处理将允许我们研究最优解对于规划中参数的微小扰动的灵敏度.

关于一般非凸规划的对偶性，可以说的事很少，因为这题目差不多正处于五十年代初凸对偶性所处的同样阶段. 这方面有少量工作，如[27, 28]，结果不是十分满意的. 某些特殊情形将在第 6 章和第 7 章中提到. 另外的参考文献在 6.2 节的末尾提及.

5.1 共轭函数

在上一章中我们定义了 R^n 上的凸函数 f 的闭包，它是具有值 $h(x)\leqslant f(x)$ 的所有线性（仿射）函数 h 的点式上确界．设 $h(x)=\xi^T x-b$，我们可以问如下问题：给定系数向量 $\xi\in R^n$，b 应当取什么值才可以保证，对一切 $x\in R^n$，$h(x)$ 小于等于 $f(x)$？显然，b 必须对一切 $x\in R^n$ 满足

$$b\geqslant\xi^T x-f(x), \tag{5.1}$$

所以

$$b\geqslant\sup_x\{\xi^T x-f(x)\}. \tag{5.2}$$

除非明显指出，否则，上确界总理解为在整个 R^n 上取．一般说满足(5.2)的 b 的最大下界是 ξ 的函数，称为 f 的共轭函数，记为 f^*．这样，

$$f^*(\xi)=\sup_x\{\xi^T x-f(x)\}. \tag{5.3}$$

共轭函数的几何解释可以在图 5.1 中看到，具有系数向量 ξ、且在函数 f 下方的最高线性函数与 f 轴的交点等于 $-f^*(\xi)$．改变 ξ 将相应地提高或降低 $-f^*(\xi)$的值，并且移动 h 切于 f 的点 x（如果它存在）．虽然共轭函数可以看作一个纯粹的数学概念，它却有明确的经济解释．例如，设制造数量 $x_1, x_2, \cdots, x_n$ 的 n 种货

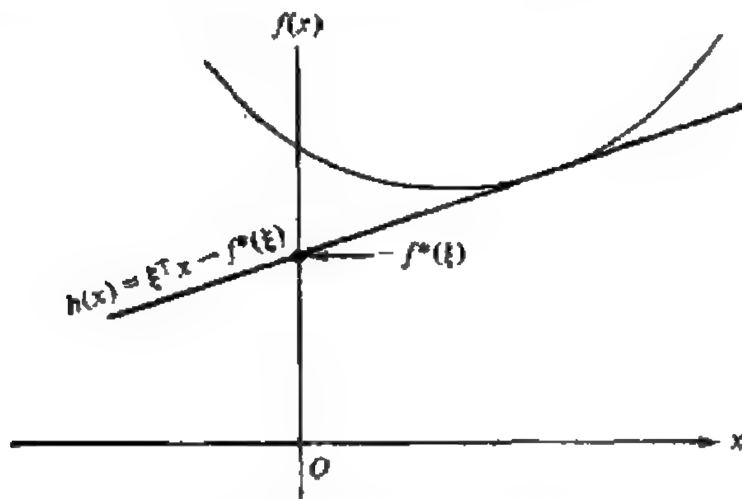


图 5.1 共轭函数的几何解释

物的费用由 $f(x)$ 给出. 这些货物分别能以价格 $\xi_1, \xi_2, \cdots, \xi_n$ 出售, 厂主的问题是选择 $x_1, \cdots, x_n$ 的值, 使得从出售这些货物所得的利润最大. 显然, 他的利润是收入 $\sum_i \xi_i x_i$ 和生产成本 $f(x)$ 的差, 如(5.3)表明的那样, 利润的最大可能值作为价格的函数由 $f^*(\xi)$ 给出.

让我们考察共轭函数的某些性质. 由(4.36), f^* 的上图象是

$$P(f^*)=\{(\xi, \alpha):\xi\in R^n,\ \alpha\in R,\ \sup_x\{\xi^T x-f(x)\}\leqslant\alpha\}, \qquad (5.4)$$

或

$$P(f^*)=\{(\xi, \alpha):\xi\in R^n,\ \alpha\in R,\ \xi^T x-\alpha\leqslant f(x),\ \forall x\in R^n\}. \qquad (5.5)$$

将上式与(4.63)比较, 可以断定: $P(f^*)$恰是支撑集 $L(f)$. 读者容易验证 $P(f^*)$是凸集, 所以 f^* 是一个凸函数. 让我们计算 f^{**}, 即 f^* 的共轭函数. 由(5.3)有

$$f^{**}(x)=\sup_\xi\{\xi^T x-f^*(\xi)\} \qquad (5.6)$$

$$=\sup_{\{(\xi,\alpha^*):f^*(\xi)\leqslant\alpha^*\}}\{\xi^T x-\alpha^*\} \qquad (5.7)$$

$$=\sup_{\{(\xi,\alpha^*):\xi^T y-\alpha^*\leqslant f(y),\ \forall y\in R^n\}}\{\xi^T x-\alpha^*\} \qquad (5.8)$$

$$=\sup_{(\xi,\alpha^*)\in L(f)}\{\xi^T x-\alpha^*\} \qquad (5.9)$$

$$=\mathrm{cl}\, f(x). \qquad (5.10)$$

我们将这些结果概括为下面的定理.

定理 5.1

设 f 是 R^n 上的凸函数, 则共轭函数 f^* 也是凸函数, 且 $P(f^*)=L(f)$. 此外 $f^{**}=\mathrm{cl}\, f$.

容易说明 $L(f^*)=P(f^{**})=P(\mathrm{cl}\, f)$和 $\mathrm{cl}\, f^*=(\mathrm{cl}\, f)^*$. 因为对每个 $x\in R^n$, $\mathrm{cl}\, f(x)\leqslant f(x)$, 我们有

$$\xi^T x-\mathrm{cl}\, f(x)\geqslant\xi^T x-f(x) \qquad (5.11)$$

以及$(\mathrm{cl}\, f)^*=\mathrm{cl}\, f^*\geqslant f^*$. 但也有 $\mathrm{cl}\, f^*\leqslant f^*$, 所以 $\mathrm{cl}\, f^*=f^*$, f^* 确是一个闭凸函数.

闭的非正常凸函数只能是, 对一切 $x\in R^n$, 或者 $f(x)=+\infty$,

或者 $f(x)=-\infty$. 前者对一切 $\xi\in R^n$, $f^*(\xi)=-\infty$; 后者 $f^*(\xi)=+\infty$. 这样,我们建立了下列定理.

定理 5.2

共轭函数 f^* 是一个闭凸函数. 它是正常凸函数的充要条件为 f 是正常凸的.

设 f 是 R^n 上的凸函数, α 是实数, z 是 R^n 中的向量, 读者容易证明共轭函数的下列基本性质:

(i) 若 $\phi(x)=f(x)+\alpha$, 则 $\phi^*(\xi)=f^*(\xi)-\alpha$. (5.12)

(ii) 若 $\phi(x)=f(x+z)$, 则 $\phi^*(\xi)=f^*(\xi)-\xi^Tz$. (5.13)

(iii) 若 $\phi(x)=f(\alpha x)$, $\alpha\neq0$, 则 $\phi^*(\xi)=f^*\left(\dfrac{\xi}{\alpha}\right)$. (5.14)

(iv) 若 $\phi(x)=\alpha f(x)$, $\alpha>0$ 则 $\phi^*(\xi)=\alpha f^*\left(\dfrac{\xi}{\alpha}\right)$. (5.15)

例 5.1.1

考虑几对简单的共轭函数.

(a) 设 $\phi(x)=a^Tx-b$, $x\in R^n$. 首先计算 $f(x)=a^Tx$ 的共轭函数,

$$f^*(\xi)=\sup_x\{\xi^Tx-a^Tx\}=\begin{cases}0, & \xi=a,\\ +\infty, & \xi\neq a,\end{cases} \tag{5.16}$$

所以由(5.12)有

$$\phi^*(\xi)=\begin{cases}b, & \xi=a,\\ +\infty, & \xi\neq a.\end{cases}$$

(b) 类似于(a), 我们有下面一对共轭函数. 设当 $x\in C\subset R^n$ 时 $\phi(x)=0$, 否则 $\phi(x)=+\infty$, 其中 C 是一个给定的凸集. 这函数称为 C 的指示函数, 它的共轭是

$$\phi^*(\xi)=\sup_{x\in C}\xi^Tx, \tag{5.17}$$

ϕ^* 称为 C 的支撑函数.

注意, 空集的指示函数恒为 $+\infty$, 它的共轭, 即支撑函数恒为 $-\infty$. 也要注意, 支撑函数是正齐次的, 它是正常凸函数的充要条件为 C 非空. 反之, 请读者证明, 给了一个正齐次的正常凸函数, 它

的共轭总是一个指示函数.

(c) 设 $\phi(x)=\frac{1}{2}x^Tx$, $f(x)=x^Tx$, $x\in R^n$, 则

$$f^*(\xi)=\sup_x\{\xi^Tx-x^Tx\}. \tag{5.18}$$

用简单的微分法, 我们可以看到上确界在 $x=\frac{1}{2}\xi$ 处达到, 这样 $f^*(\xi)=\frac{1}{4}\xi^T\xi$, 由(5.15), $\phi^*(\xi)=\frac{1}{2}\xi^T\xi$.

(d) 设 $x\in R$, $\phi(x)=\begin{cases}-\log x, & x>0,\\ +\infty, & \text{其他},\end{cases}$

则 $\phi^*(\xi)=\sup\limits_{x>0}\{\xi x+\log x\}$. 对 $\xi\geqslant 0$, 因为 $x\to+\infty$ 时 $\xi x+\log x$ 可任意大, 我们有 $\phi^*(\xi)=+\infty$; 而对 $\xi<0$, 用微分法可求得其上确界, 从而 $\phi^*(\xi)=-1-\log(-\xi)$. 这样

$$\phi^*(\xi)=\begin{cases}-1-\log(-\xi), & \xi<0,\\ +\infty, & \text{其他},\end{cases} \tag{5.19}$$

希望读者去求上述四种情形中的共轭函数的共轭. ▌

共轭函数也可看作是经典的 Legendre 变换的推广[4,17]. 设 f 是具有有效区域 $\mathrm{ED}(f)\subset R^n$ 的闭凸函数, 它在 $\mathrm{ED}(f)$ 的非空内部 C 上可微. 进一步设 f 的次微分 ∂f 是一对一映射. 这等价于条件: f 是严格凸的, 且在它的有效区域的边界点附近变得无限陡[80].

在前面关于 f 的假定下, 在每一点 $x\in C$ 处, 次微分 ∂f 含有单个向量 $\xi=\nabla f(x)$. 如果 ∇f 是一对一的, 则它有逆函数 ∇f^{-1}, 其定义域是 ∇f 的值域 D, 即梯度映射下 C 的象. 函数 ∇f^{-1} 也是一对一的[9]. 定义

$$H(\xi)=\xi^T\nabla f^{-1}(\xi)-f(\nabla f^{-1}(\xi)). \tag{5.20}$$

则 H 也是 D 上的可微凸函数且 $\nabla H=\nabla f^{-1}$. 从 $(C, f(x))$ 到 $(D, H(\xi))$ 的变换称为 **Legendre 变换**.

设 $\xi\in D\subset R^n$, 由(4.110), 对一切 $y\in R^n$ 有

$$\xi^Tx-f(x)\geqslant\xi^Ty-f(y), \tag{5.21}$$

这样,

$$\xi^T x - f(x) = \sup_y\{\xi^T y - f(y)\} = f^*(\xi). \tag{5.22}$$

因为 $x = \nabla f^{-1}(\xi)$，所以 $H(\xi)$ 与 $f^*(\xi)$ 在 D 上一致．在关于 ∇f 的同样假定下和 $C = R^n$ 这一特殊情况中，f 是严格凸函数，且对一切 $x \neq 0$ 成立

$$\lim_{\lambda \to \infty} \frac{f(\lambda x)}{\lambda} = +\infty, \tag{5.23}$$

于是 $D = R^n$，且对一切 $\xi \in R^n$ 有 $H(\xi) = f^*(\xi)$[30]．

所以共轭函数是 Legendre 变换对凸函数的扩充，这些凸函数在扩充的实数集上取值，而不必连续或可微．

在次梯度和共轭函数之间也有密切联系，这可从下面结果中看到．

定理 5.3

设 f 是 R^n 上的凸函数，则 $\xi \in \partial f(x)$ 的充要条件为

$$f^*(\xi) = \xi^T x - f(x).$$

【证明】 由(4.110)，$\xi \in \partial f(x)$ 的充要条件为(5.21)成立，也就是(5.22)成立． ▌

定理 5.4

设 f 是 R^n 上的凸函数，若 $\xi \in \partial f(x)$，则 $x \in \partial f^*(\xi)$．若 f 又在 x 点是闭的且 $x \in \partial f^*(\xi)$，则 $\xi \in \partial f(x)$．

【证明】 由前面的定理，当且仅当 $f^*(\xi) = \xi^T x - f(x)$ 时 $\xi \in \partial f(x)$；当且仅当 $f^{**}(x) = \xi^T x - f^*(\xi)$ 时 $x \in \partial f^*(\xi)$．在练习 4.N 中已要读者证明：若在 $\hat{x}$ 处 $\partial f(\hat{x})$ 非空，则 $\mathrm{cl}\, f(\hat{x}) = f(\hat{x})$．所以 $\xi \in \partial f(x)$ 蕴涵

$$\mathrm{cl}\, f(x) = f^{**}(x) = f(x) = \xi^T x - f^*(\xi), \tag{5.24}$$

从而 $x \in \partial f^*(\xi)$．反之，设 f 是闭的且 $x \in \partial f^*(\xi)$，则(5.24)成立，即

$$f^*(\xi) = \xi^T x - f(x), \tag{5.25}$$

从而 $\xi \in \partial f(x)$． ▌

让我们简短地讨论凹函数的共轭．一般地说，在凸函数的共轭函数理论中的每个结果，可以对凹函数类似地建立，只要想到，

函数 g 是凹的充要条件为 $f=-g$ 是凸的. 特别地,凹函数 g 的共轭函数 g^* 定义为

$$g^*(\xi)=\inf_x\{\xi^T x-g(x)\}, \tag{5.26}$$

这里 g^* 是闭凹函数, 它的共轭是 $\mathrm{cl}\, g$, 其闭包运算的定义类似于凸的情形. 但是当 $f=-g$ 时, g^* 与 $-f^*$ 并不一致.

事实上, 有

$$f^*(\xi)=\sup_x\{\xi^T x-f(x)\} \tag{5.27}$$

$$=\sup_x\{-[-\xi^T x-g(x)]\} \tag{5.28}$$

$$=-\inf_x\{-\xi^T x-g(x)\}=-g^*(-\xi). \tag{5.29}$$

最后我们注意,对既非凸也非凹的函数,共轭函数可以象上面那样来定义. 设 f 是取值在扩充的实数集上且在 R^n 上定义的任意函数, 则它的**凸共轭** f^* 由(5.3)定义, 这是一个闭凸函数, 且是满足 $f_0\leqslant f$ 的最大凸函数 f_0 的闭包的共轭. 换言之, f^* 是一个凸函数的闭包的共轭, 这凸函数是 $P(f)$ 的凸包的下部轮廓(见定理4.14). 对这样的任意函数, 可以用明显类似的方式定义凹共轭.

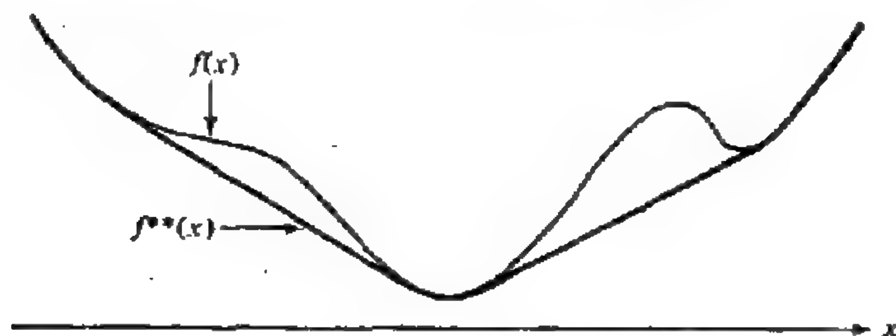


图 5.2 非凸函数的凸共轭的共轭

正如下节将表明的凸函数和凹函数的共轭在非线性凸规划中起重要作用那样, 扩充到特殊类型的非凸规划问题的对偶性关系已借助某些非凸函数的共轭而得到(将在下两章讨论)。

5.2 对偶凸规划

正如前面已指出的, 我们将在某种广泛意义下建立非线性凸

规划的对偶性关系，它将使我们深刻地看到最优值对于规划中参数的某种扰动的灵敏度．本节我们将处理前一章引进的(CP)的下列形式：

(SP) $$\min_x f(x) \tag{5.30}$$

受限制于

$$g_i(x)\geqslant 0,\quad i=1,\ \cdots,\ m, \tag{5.31}$$

其中 f 是 R^n 上的正常凸函数，g_i 是 R^n 上的正常凹函数．我们称规划(SP)为**标准原有凸规划问题**．当 f 和 g_i 是实值时，Geoffrion[18] 称之为"典则的"原有问题，Rockafellar[32] 称之为"普通的"凸规划．它与第4章表述的(CP)不同处在于，或者不出现线性等式约束，或者它们作为两个不等式约束出现．前面提到的扰动，例如可以在约束的右边引进．也就是，(5.31)中的不等式可代之以更一般的约束

$$g_i(x)\geqslant u_i,\quad i=1,\ \cdots,\ m; \tag{5.32}$$

其中 u_i 是实参数．以(5.32)代替(5.31)作为约束的(SP)的最优解可以看作 u 的函数，而对原点附近的向量 u，考查最优解的性态是有意义的．包含一大类扰动的更一般的规划也由 Rockafellar[32] 研究过．这规划可写成

(QP) $$\min f(x-y^0) \tag{5.33}$$

受限制于

$$g_i(x-y^i)\geqslant u_i,\quad i=1,\ \cdots,\ m, \tag{5.34}$$

这里 $y^i\in R^n$，$i=0,\ 1,\ \cdots,\ m$，是扰动，它们平移 f 与 g_i，u_i 如以前那样是约束右端的扰动．规划(SP)显然是(QP)当 $y^i=0$ 和 $u_i=0$ 时的特殊情形．

用 C 记 f 的有效区域．我们用下面关系式定义 R^n 上的一个新的凸函数 ϕ：

$$\phi(x)=\begin{cases} f(x),\ x\in C\ 且\ g_i(x)\geqslant 0,\ i=1,\ \cdots,\ m, \\ +\infty,\ 其他, \end{cases} \tag{5.35}$$

(SP)就可等价地写成 ϕ 的(无约束)极小化问题．如果我们用 $w\in R^k$ 记某个扰动向量，就可将(5.35)推广成

$$\phi(x, w)=\begin{cases} f(x, w), (x, w)\in \mathrm{ED}(f)\subset R^{n+k} \\ \quad 且 g_i(x, w)\geqslant 0, i=1, \cdots, m, \\ +\infty, \quad 其他, \end{cases} \tag{5.36}$$

选择一个将 w 合并入所含函数的适当形式，(SP)就变成在所有 $x\in R^n$ 上 $\phi(x,0)$ 的极小化问题. (5.36)给出的公式推动了现有的一般分析，这分析是基于 Geoffrion[18], Hamala[19], Rockafellar[31, 32] 的工作.

设 ϕ 是两个向量 $x\in R^n$ 和 $w\in R^k$ 的正常凸函数，ϕ 的共轭函数定义为

$$\phi^*(\xi, \lambda)=\sup_{x,w}\{\xi^T x+\lambda^T w-\phi(x, w)\}, \tag{5.37}$$

所以对一切 $x\in R^n$, $\xi\in R^n$, $w\in R^k$, $\lambda\in R^k$, 我们有

$$\phi(x, w)+\phi^*(\xi, \lambda)\geqslant \xi^T x+\lambda^T w. \tag{5.38}$$

特别地，令 $w=0$, $\xi=0$, 对每个 $x\in R^n$ 和 $\lambda\in R^k$, 我们得到

$$\phi(x, 0)+\phi^*(0, \lambda)\geqslant 0. \tag{5.39}$$

作为上述关系式的结果，定义下列一对凸规划:

$$(\mathrm{P}_\phi) \qquad \min_x \phi(x, 0), \quad (原有规划) \tag{5.40}$$

$$(\mathrm{D}_{\phi^*}) \qquad \max_\lambda -\phi^*(0, \lambda). \quad (对偶规划) \tag{5.41}$$

需要指出，在这两个规划中我们寻找那样的 x 和 λ, 它们分别给出 $\phi(x, 0)$ 的极小值和 $-\phi^*(0, \lambda)$ 的极大值，但在很多情况下没有这样的 x 和 λ 存在，虽然 $\phi(x, 0)$ 可能有有限的下确界，$-\phi^*(0, \lambda)$ 有有限的上确界. 我们分别称 $\phi(x, 0)$ 的下确界和 $-\phi^*(0, \lambda)$ 的上确界为(P_ϕ)和(D_{ϕ^*})的最优值. 下面的结果是(5.39)的直接结论.

定理 5.5(弱对偶性定理)

$$\inf_x \phi(x, 0)\geqslant \sup_\lambda -\phi^*(0, \lambda). \tag{5.42}$$

推论 5.6

若

$$\phi(x^*, 0)=-\phi^*(0, \lambda^*), \tag{5.43}$$

则 x^* 和 λ^* 分别是(P_ϕ)和(D_{ϕ^*})的最优解.

若存在 $\hat{x}$ 使得 $\phi(\hat{x},0)<+\infty$ 或者等价地，若 $(\hat{x},0)\in \mathrm{ED}(\phi)$，则说规划$(\mathrm{P}_\phi)$是相容的. 类似地，若存在 $\hat{\lambda}$ 使得 $\phi^*(0,\hat{\lambda})<+\infty$，则说规划$(\mathrm{D}_{\phi^*})$是相容的. 若对一切 $x\in R^n$，$(x, 0)\overline{\in}\mathrm{ED}(\phi)$，则说$(\mathrm{P}_\phi)$是不相容的. 若对一切 $\lambda\in R^k$，$(0, \lambda)\overline{\in}\mathrm{ED}(\phi^*)$，则说$(\mathrm{D}_{\phi^*})$是不相容的. 用这术语，从定理 5.5 可得下面结果.

推论 5.7

若

$$\inf_x \phi(x, 0)=-\infty, \tag{5.44}$$

则(D_{ϕ^*})是不相容的. 若

$$\sup_\lambda -\phi^*(0, \lambda)=+\infty, \tag{5.45}$$

则(P_ϕ)是不相容的.

在现有结果中，凸性起着核心作用，有趣的是要注意：对不一定是凸函数的 ϕ，关系式(5.37)至(5.43)也是成立的，只要避免无意义的求和 $+\infty+(-\infty)$. 例如，如果 $\phi(x, 0)$在 R^n 上有有限值，就可保证这点. 对偶规划的任何能行解 λ 提供 $-\phi^*(0, \lambda)$的一个值，可作为(P_ϕ)的最优解的下界. 但是总希望有较(5.42)更强的结果——(P_ϕ) 和 (D_{ϕ^*}) 的最优解正好相等. 这里是凸规划和非凸规划的主要差别之一：对于凸的 ϕ，在适当的正则性假定下，(5.42)作为等式成立；而当 ϕ 非凸时，可能有一个对偶性间隙，它就是当(5.42)作为严格不等式成立时 (P_ϕ) 和 (D_{ϕ^*}) 的最优值之差.

现在设 ϕ 是 R^{n+k} 上的闭正常凸函数，则 $\phi^{**}=\phi$. 把(D_{ϕ^*})重新表述为原有规划

$$(\mathrm{P}_{\phi^*}) \qquad \min_\lambda \phi^*(0, \lambda), \tag{5.46}$$

我们得到对偶规划

$$(\mathrm{D}_\phi) \qquad \max_x -\phi^{**}(x, 0)=\max_x -\phi(x, 0), \tag{5.47}$$

它在下述意义下等价于本来的原有规划(5.40)：向量 x 是(P_ϕ)的最优解，当且仅当它是(D_ϕ)的最优解. 这样，对闭正常凸函数，对

偶性关系是对称的:(D_{ϕ^*})的对偶是(P_ϕ).

与ϕ和ϕ^*相联系,我们定义原有和对偶扰动函数Φ和Ψ如下:

$$\Phi(w)=\inf_x \phi(x,\ w), \tag{5.48}$$

$$\Psi(\xi)=\inf_\lambda \phi^*(\xi,\ \lambda). \tag{5.49}$$

下面是扰动函数的一个重要性质.

定理5.8

设ϕ是R^{n+k}上的正常凸函数,则扰动函数Φ是R^k上的凸函数.

【证明】 我们必须说明上图象$P(\Phi)\subset R^{k+1}$是凸集. 因为ϕ是正常凸的,所以$P(\phi)$是非空的,从而$P(\Phi)$也非空. 设$(w^1,\ \alpha^1)\in P(\Phi)$, $(w^2,\ \alpha^2)\in P(\Phi)$. 可知,对一切$\varepsilon>0$,能找到向量$x_1,\ x_2$,使得

$$\phi(x^1,\ w^1)<\inf_x \phi(x,\ w^1)+\varepsilon=\Phi(w^1)+\varepsilon\leqslant\alpha^1+\varepsilon, \tag{5.50}$$

$$\phi(x^2,\ w^2)<\inf_x \phi(x,\ w^2)+\varepsilon=\Phi(w^2)+\varepsilon\leqslant\alpha^2+\varepsilon. \tag{5.51}$$

因为$P(\phi)$是凸集,对一切$q_1\geqslant0,\ q_2\geqslant0,\ q_1+q_2=1$和$\varepsilon>0$,我们有

$$\phi(q_1x^1+q_2x^2,\ q_1w^1+q_2w^2)<q_1\alpha^1+q_2\alpha^2+\varepsilon, \tag{5.52}$$

所以

$$\Phi(q_1w^1+q_2w^2)=\inf_x \phi(x,\ q_1w^1+q_2w^2)$$

$$\leqslant\phi(q_1x^1+q_2x^2,\ q_1w^1+q_2w^2). \tag{5.53}$$

结合(5.52)和(5.53),可得$\Phi(q_1w^1+q_2w^2)\leqslant q_1\alpha^1+q_2\alpha^2$,因此$P(\Phi)$是凸集. 】

容易证明,Φ的有效区域由下式给出:

$$\mathrm{ED}(\Phi)=\{w: w\in R^k,\ \phi(x,\ w)<+\infty,\ 对某\ x\}. \tag{5.54}$$

注意,用类似于上面的论证可得$\Psi(\xi)$是凸函数.

在第4章中定义的凸函数的次梯度和方向导数在对偶性关系中起重要作用. 我们从一个以后将用到的关于次梯度存在的准则

开始.

引理 5.9

设 f 是凸函数，$x\in R^n$ 是使 $f(x)$ 有限的一点，则 f 在 x 有次梯度即 $\partial f(x)\neq\emptyset$ 的充要条件是：对一切 $y\in R^n$，

$$D^+f(x;\ y)>-\infty. \tag{5.55}$$

【证明】 实际上将证明相反的方面，即 $\partial f(x)=\emptyset$ 的充要条件是：对某个 $y\in R^n$，

$$D^+f(x;\ y)=D^-f(x;\ y)=-\infty. \tag{5.56}$$

设 $\hat{x}$ 是使 $f(\hat{x})$ 有限的一点．由定理 4.21，$d(y)=D^+f(\hat{x};\ y)$ 是正齐次凸函数，它的共轭 d^* 是闭凸指示函数（见练习 5.C），这样就存在凸集 E，使得

$$d^*(\eta)=\begin{cases}0, & \eta\in E,\\ +\infty, & \eta\bar{\in} E.\end{cases} \tag{5.57}$$

对一切 $\eta\in E$，我们有

$$d^*(\eta)=\sup_y\{\eta^Ty-d(y)\}=0. \tag{5.58}$$

所以集 E 由下式给出：

$$E=\{\eta:\forall y,\ D^+f(\hat{x};\ y)\geqslant\eta^Ty\}. \tag{5.59}$$

由定理 4.22 可推得集 E 是 f 在 $\hat{x}$ 处所有次梯度之集，即 $E=\partial f(\hat{x})$．d^* 的共轭是 $\partial f(\hat{x})$ 的支撑函数，它是 $d^{**}=\mathrm{cl}\,d=\mathrm{cl}\,D^+f(x;\ y)$．空集的指示函数恒为 $+\infty$，它的共轭，即支撑函数恒为 $-\infty$．这样，当且仅当对一切 y，$\mathrm{cl}\,D^+f(\hat{x};\ y)=-\infty$ 时，$\partial f(\hat{x})=\emptyset$．也就是对某 y，$D^+f(\hat{x};\ y)$ 必须取值 $-\infty$． ∎

考虑凸扰动函数 Φ．原有规划称为稳定的，如果：(a) $\Phi(0)$ 是有限的，且没有方向 y 使得 $D^+\Phi(0;\ y)=-\infty$，或者 (b) $\Phi(0)=-\infty$．换言之，(P_ϕ) 是稳定的，如果 $\inf\limits_x\phi(x,\ 0)$ 有限且扰动函数 $\Phi(w)$ 对 0 的邻域中任何 w 不是无限陡地减小，或者(P_ϕ)的最优值无下界．作为引理 5.9 的结果，我们可以给稳定性以另一特征：规划(P_ϕ)是稳定的，当且仅当 $\partial\Phi(0)$ 是非空集．注意如果 $\Phi(0)=-\infty$，则 $\partial\Phi(0)=R^k$．非线性规划的稳定性在没有凸性的假定下，

曾由 Evans 和Gould[12]研究过，而对某一类凸规划，曾由Dantzig、Folkman、Shapiro[7] 研究过.

例 5.2.1

让我们考虑稳定规划和不稳定规划的一些例子.

(a) 令

$$\phi_1(x,\ w)=\begin{cases} x, & x \geqslant w, \\ +\infty, & x<w, \end{cases} \tag{5.60}$$

其中 $x\in R$, $w\in R$. 容易证明，对一切 $w\in R$,

$$\Phi_1(w)=\inf_x \phi_1(x,\ w)=w, \tag{5.61}$$

且 $\Phi_1(0)=0$, 所以 Φ_1 在 0 处有限. 确实，$\partial\Phi_1(0)$由下面一个点组成:

$$\partial\Phi_1(0)=\nabla\Phi_1(0)=1, \tag{5.62}$$

因而(P_{ϕ_1})是稳定的.

(b) 令

$$\phi_2(x,\ w)=\begin{cases} -x, & x\geqslant w, \\ +\infty, & x<w. \end{cases} \tag{5.63}$$

对一切 $x\in w$, 我们可得

$$\Phi_2(w)=\inf_x \phi_2(x,\ w)=-\infty, \tag{5.64}$$

所以 $\Phi_2(0)=-\infty$, 从而(P_{ϕ_2})又是稳定的.

(c) 取

$$\phi_3(x,\ w)=\begin{cases} x, & (x)^2\leqslant w, \\ +\infty, & (x)^2>w. \end{cases} \tag{5.65}$$

则

$$\Phi_3(w)=\begin{cases} -\sqrt{w}, & w\geqslant 0, \\ +\infty, & w<0. \end{cases} \tag{5.66}$$

这里 Φ_3 是闭正常凸的，如图 5.3 所示. 在 $w=0$ 有 $\Phi_3(0)=0$, 但

$$D^+\Phi_3(0;1)=\lim_{w\to 0^+}\frac{\Phi_3(w)-\Phi_3(0)}{w}=\lim_{w\to 0^+}\frac{-\sqrt{w}}{w}=-\infty, \tag{5.67}$$

这样，(P_{ϕ_3})是不稳定的. ∎

稳定性是加在(P_ϕ)上的一个条件，我们将在下面的定理中用这条件以建立(P_ϕ)和(D_{ϕ^*})的最优值的等式，建立对偶规划(D_{ϕ^*})的最优解的存在性，并建立以Φ的次梯度表示的这样一个解的特征。首先我们需要下面的引理。

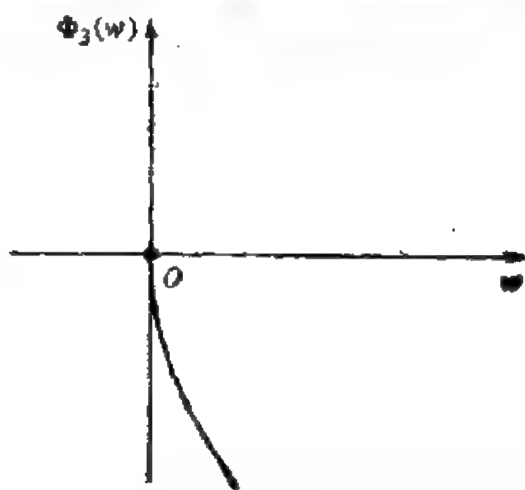


图 5.3 不稳定规划的扰动函数

引理 5.10

设$\phi(x,w)$是R^{n+k}上的正常凸函数，定义$\Phi(w)$如(5.48)，则

$$\phi^*(0, \lambda)=\Phi^*(\lambda). \quad (5.68)$$

【证明】 我们有

$$\phi^*(0, \lambda)=\sup_{x,w}\{0^Tx+\lambda^Tw-\phi(x, w)\} \quad (5.69)$$

$$=\sup_{w}\{\lambda^Tw-\inf_x \phi(x, w)\} \quad (5.70)$$

$$=\sup_{w}\{\lambda^Tw-\Phi(w)\}=\Phi^*(\lambda). \quad (5.71)$$

∎

我们现在叙述并证明一个基于稳定性的强对偶性定理.

定理 5.11(强对偶性定理)

设ϕ是R^{n+k}上的正常凸函数，并设$\Phi(0)$有限，则规划(D_{ϕ^*})有最优解λ^*且

$$\Phi(0)=\inf_x \phi(x, 0)=\max_\lambda -\phi^*(0, \lambda)=-\phi^*(0, \lambda^*) \quad (5.72)$$

的充要条件为(P_ϕ)是稳定的. 此外，当且仅当(5.72)成立时，λ^*是$\Phi(0)$的次梯度.

【证明】 规划(P_ϕ)是稳定的充要条件为$\partial\Phi(0)$非空. 由次梯度的定义，$\lambda^*\in\partial\Phi(0)$的充要条件为：对一切$w\in R^k$成立

$$\Phi(w)\geqslant\Phi(0)+\lambda^{*T}w. \quad (5.73)$$

所以,

$$\Phi(0)\geqslant\inf_w\{\Phi(w)-\lambda^{*T}w\}=-\Phi^*(\lambda^*)\geqslant\Phi(0). \quad (5.74)$$

由引理 5.10 可知,

$$-\Phi^*(\lambda^*)=-\phi^*(0,\ \lambda^*)=\Phi(0)=\inf_{x}\phi(x,\ 0); \tag{5.75}$$

由定理 5.5, 最后可得

$$-\phi^*(0,\ \lambda^*)=\max_{\lambda}-\phi^*(0,\ \lambda)=\inf_{x}\phi(x,\ 0). \tag{5.76}$$

∎

与(P_ϕ)的稳定性定义完全相类似, 我们也可以定义$(D_{\phi*})$的稳定性: 规划$(D_{\phi*})$是稳定的, 当且仅当$\partial\Psi(0)\neq\emptyset$. 如果假定$\phi(x,\ w)$是闭正常凸函数, 那么可得到类似于引理 5.10 和定理 5.11 的结果.

引理 5.12

设$\phi(x,\ w)$是R^{n+m}上的闭正常凸函数,则

$$\phi(x,\ 0)=\Psi^*(x). \tag{5.77}$$

【证明】 因为ϕ是闭的凸函数,

$$\phi(x,\ 0)=\phi^{**}(x,\ 0)=\sup_{\xi,\lambda}\{\xi^Tx+\lambda^T0-\phi^*(\xi,\ \lambda)\} \tag{5.78}$$

$$=\sup_{\xi}\{\xi^Tx-\inf_{\lambda}\phi^*(\xi,\ \lambda)\} \tag{5.79}$$

$$=\sup_{\xi}\{\xi^Tx-\Psi(\xi)\}=\Psi^*(x). \tag{5.80}$$

∎

这样我们就有如下对偶结果.

定理 5.13

设ϕ是闭正常凸函数, 且设$\Psi(0)$有限, 则规划(P_ϕ)有最优解x^*且

$$\phi(x^*,\ 0)=\min_{x}\phi(x,\ 0)=\sup_{\lambda}-\phi^*(0,\ \lambda) \tag{5.81}$$

的充要条件为$(D_{\phi*})$是稳定的. 此外, 当且仅当(5.81)成立时, x^*是$\Psi(0)$的次梯度.

【证明】 类似于定理 5.11 的证明. ∎

下面两个推论是关于ϕ为闭正常凸函数时(P_ϕ)和$(D_{\phi*})$的最优解的存在性的.

推论 5.14

设ϕ是R^{n+k}上的闭正常凸函数，并设$\Phi(0)$有限，若(P_ϕ)和(D_{ϕ^*})都是稳定的，则两者都有最优解，且

$$+\infty>\min_x \phi(x,\ 0)=\max_\lambda -\phi^*(0,\ \lambda)>-\infty. \qquad (5.82)$$

【证明】 由定理5.11和5.13，若(P_ϕ)和(D_{ϕ^*})都是稳定的，则

$$\min_x \phi(x,\ 0)=\max_\lambda -\phi^*(0,\ \lambda), \qquad (5.83)$$

因为ϕ是正常凸的，$\phi(x,\ 0)$达到的极小值不能是$-\infty$. 类似地，$-\phi^*(0,\ \lambda)$的极大值不能是$+\infty$. ∎

推论5.15

设ϕ是闭正常凸函数，则(P_ϕ)是稳定的且有最优解x^*的充要条件为(D_{ϕ^*})是稳定的且有最优解λ^*.

【证明】 由定理5.11和5.13，规划(P_ϕ)是稳定的且有最优解x^*的充要条件为

$$\phi(x^*,0)=\min_x \phi(x,0)=\max_\lambda -\phi^*(0,\lambda)=-\phi^*(0,\ \lambda^*), \qquad (5.84)$$

但这等价于条件：规划(D_{ϕ^*})是稳定的且有解λ^*. ∎

例5.2.2

重新考虑前面例子中的函数.

(a) 首先设

$$\phi_1(x,\ w)=\begin{cases} x, & x\geqslant w, \\ +\infty, & x<w, \end{cases} \qquad (5.85)$$

则

$$\phi_1^*(\xi,\ \lambda)=\begin{cases} 0, & \xi+\lambda=1,\ \lambda\geqslant 0, \\ +\infty, & 其他, \end{cases} \qquad (5.86)$$

所以稳定规划(P_{ϕ_1})可以写成

$$\min_x \phi_1(x,\ 0)=\min_{x\geqslant 0} x, \qquad (5.87)$$

而$(D_{\phi_1^*})$由下式给出：

$$\max_\lambda -\phi^*(0,\ \lambda)=\max_{\lambda=1} 0. \qquad (5.88)$$

显然，$\min_{x\geqslant 0} x=0$，且(P_{ϕ_1})和$(D_{\phi_1^*})$的最优解分别是$x^*=0$,

$\lambda^*=1$. 我们已经看到，对一切 $w\in R$ 有 $\Phi_1(w)=w$，且 $\Phi_1(0)=0$，$\partial\Phi_1(0)=\{1\}$.

类似地，

$$\Psi_1(\xi)=\begin{cases}0, & \xi\leqslant 1,\\ +\infty, & \xi>1,\end{cases} \tag{5.89}$$

由此推得 $\Psi_1(0)=0$，且 $\partial\Psi_1(0)=\{0\}$，所以 $(D_{\phi_1^*})$ 也是稳定的. 此外，我们有 $x^*\in\partial\Psi(0)$ 和 $\lambda^*\in\partial\Phi(0)$.

(b) 其次，取

$$\phi_2(x, w)=\begin{cases}-x, & x\geqslant w,\\ +\infty, & x<w,\end{cases} \tag{5.90}$$

则

$$\phi_2^*(\xi, \lambda)=\begin{cases}0, & \xi+\lambda=-1,\ \lambda\geqslant 0,\\ +\infty, & \text{其他},\end{cases} \tag{5.91}$$

(P_{ϕ_2}) 的最优值为 $-\infty$，也就是

$$\inf_x \phi_2(x, 0)=\inf_{x\geqslant 0}-x=-\infty. \tag{5.92}$$

现在转向对偶问题 $(D_{\phi_2^*})$. 我们可以看到，对一切实数 λ，$(0, \lambda)\overline{\in}ED(\phi_2^*)$（因为 $\lambda=-1$, $\lambda\geqslant 0$ 无解）.

(c) 设

$$\phi_3(x, w)=\begin{cases}x, & (x)^2\leqslant w,\\ +\infty, & (x)^2>w.\end{cases} \tag{5.93}$$

在例 5.2.1 中我们已看到，规划 (P_{ϕ_3}) 是不稳定的. 计算 ϕ_3 的共轭函数，可以证明

$$\phi_3^*(\xi, \lambda)=\begin{cases}0, & \xi=1,\ \lambda=0,\\ -\dfrac{(\xi-1)^2}{4\lambda}, & \lambda<0,\\ +\infty, & \text{其他}.\end{cases} \tag{5.94}$$

(P_{ϕ_3}) 的最优解是 $x^*=0$ 且

$$\min_x \phi_3(x, 0)=0. \tag{5.95}$$

转向对偶规划 $(D_{\phi_3^*})$，我们得到

$$\sup_{\lambda<0} -\phi_3^*(0, \lambda) = \sup_{\lambda<0} \frac{1}{4\lambda} = 0. \tag{5.96}$$

但没有实数 λ^*，使上确界达到． ▮

我们已经看到，稳定性是关于凸规划对偶性的一个非常重要的性质．实际上，我们将看到它在建立凸规划的 Lagrange(Kuhn-Tucker) 乘子的存在性中起同样重要的作用．这样，研究蕴涵着稳定性的条件是重要的．在上一章我们定义了强相容规划，现在对(5.40)所给的更一般的凸规划，重新给它定义．

规划 (P_ϕ) 称为强相容的，如果存在正数 ε，使得对一切满足 $\|w\|<\varepsilon$ 的 w，可以找到 x^0 满足 $\phi(x^0, w)<+\infty$．读者可以证明，对于由(5.30)和(5.31)给出的(SP)形式的凸规划，上一章所定义的强相容性意味着存在一个 x^0，使得

$$g_i(x^0)>0, \quad i=1, \cdots, m. \tag{5.97}$$

而这条件转过来蕴涵着相应的(P_ϕ)在刚才意义下是强相容的，而(P_ϕ)是从一个在约束右边有扰动向量 w 的规划导出来的．要求读者证明下面的定理．

定理 5.16

规划(P_ϕ)是强相容的充要条件为 0 是 $\mathrm{ED}(\Phi)$ 的内点．若(P_ϕ)是强相容的，则(P_ϕ)是稳定的．

后一个断言的逆命题不真，因为有这样的稳定规划，它是相容的但不是强相容的．

我们简短地提一下对偶性间隙的概念，即原有规划的最优值严格大于相应的对偶规划的最优值的情形．在缺少稳定性的情况下，我们可以象下面说明的那样，构造这种例子．

例 5.2.8[29]

考虑极小化 $\phi(x, w)$ 的问题，其中 $x\in R^2$, $w\in R^2$ 且

$$\phi(x, w)=\begin{cases}\max[-1, -(-w_1x_2)^{1/2}], \\ \quad -w_1=x_1\geqslant 0 \text{ 且 } x_2\geqslant 0, \\ +\infty, \qquad\qquad \text{其他}.\end{cases} \tag{5.98}$$

显然，$\phi(x, 0)$或者为 0，或者为 $+\infty$，(P_ϕ)的最优值为

$$\inf_x \phi(x,\ 0) = \min_x \phi(x,\ 0) = \phi(x^*,\ 0) = 0, \tag{5.99}$$

其中 x^* 不妨取为 $x^* = (0,\ 0)$，读者可以验证

$$\Phi(w) = \begin{cases} -1, & w_1 < 0, \\ 0, & w_1 = 0, \\ +\infty, & w_1 > 0, \end{cases} \tag{5.100}$$

且 $\Phi(0) = 0$. 因为 $\partial\Phi(0)$ 是空的，或者等价地，因为 $\Phi(w)$ 在原点的每个邻域中(在负 w_1 方向上)无限陡地减小，所以规划 (P_ϕ) 是不稳定的. 读者也可以验证

$$\phi^*(0,\ \lambda) = \begin{cases} 1, & \lambda_1 \geqslant 0,\ \lambda_2 = 0, \\ +\infty, & \text{其他}, \end{cases} \tag{5.101}$$

以及 (D_{ϕ^*}) 的最优值由下式给出：

$$\sup_\lambda -\phi^*(0,\ \lambda) = \max_\lambda -\phi^*(0,\ \lambda) = -\phi^*(0,\ \lambda^*) = -1, \tag{5.102}$$

其中 λ^* 不妨取为 $\lambda^* = (0,\ 0)$. 因为

$$\min_x \phi(x,\ 0) = 0 > -1 = \max_\lambda -\phi^*(0,\ \lambda), \tag{5.103}$$

我们就得到一个对偶性间隙. ▌

在例 5.2.2(c) 中，我们有一个不具有对偶性间隙的不稳定的原有规划，而例 5.2.3 给出一个具有对偶性间隙的不稳定的原有规划. 这两个原有规划的主要差别在于相应的扰动函数：不具有对偶性间隙的一对规划的原有扰动函数是闭凸函数，而例 5.2.3 的原有扰动函数在 $w=0$ 不闭. 但注意，两个例子中的原有目标函数 ϕ 都是闭凸函数. 后面将会清楚，原有和对偶扰动函数在原点的闭性是原有和对偶规划最优值相等的充要条件.

在证明这结果之前，我们需要下面的引理.

引理 5.17

设 $\phi(x,\ w)$ 是 R^{n+k} 上的闭正常凸函数，则

$$\Phi(0) = -\mathrm{cl}\,\Psi(0), \tag{5.104}$$

$$\Psi(0) = -\mathrm{cl}\,\Phi(0). \tag{5.105}$$

【证明】 我们将证明(5.104)，要求读者证明(5.105). 我们有

$$-\mathrm{cl}\,\Psi(0) = -\Psi^{**}(0) = -\sup_x\{0^T x - \Psi^*(x)\} = \inf_x \Psi^*(x). \tag{5.106}$$

由引理 5.12 得到

$$\inf_x \Psi^*(x) = \inf_x \phi(x, 0) = \Phi(0). \tag{5.107}$$

∎

现在我们可以叙述和证明下列定理.

定理 5.18

设 $\phi(x, w)$ 是 R^{n+k} 上的闭正常凸函数，则下列陈述是等价的:

(i) $\Phi(0) = \inf_x \phi(x, 0) = \sup_\lambda - \phi^*(0, \lambda) = -\Psi(0)$.

$$\tag{5.108}$$

(ii) 原有扰动函数 Φ 在 $w=0$ 处是闭的.

(iii) 对偶扰动函数 Ψ 在 $\xi=0$ 处是闭的.

【证明】 从引理 5.17 和闭性的定义我们有，$\Phi(0) = -\Psi(0)$ 的充要条件为 $\Phi(0) = \mathrm{cl}\,\Phi(0)$. 类似地, $\Phi(0) = -\Psi(0)$ 的充要条件为 $\Psi(0) = \mathrm{cl}\,\Psi(0)$. ∎

Rockafellar[31] 称规划 (P_ϕ) 是正规的, 如果定理 5.18 的第(ii)部分成立. 类似地, (D_{ϕ^*}) 称为正规的, 如果第(iii)部分成立. 用这些定义, 我们得到上面定理的下述推论.

推论 5.19

设 $\phi(x, w)$ 是 R^{n+k} 上的闭正常凸函数. 规划 (P_ϕ) 为正规且它的最优值有限的充要条件是: (D_{ϕ^*}) 为正规且它的最优值有限, 这时它们的最优值相等. 反之, 如果 $\inf_x \phi(x, 0) = \sup_\lambda - \phi^*(0, \lambda)$, 则 (P_ϕ) 和 (D_{ϕ^*}) 两者都为正规的.

5.3 最优性条件和 Lagrange 乘子

在本节中, 我们试图将上一节导出的凸规划的对偶性关系与第 3 章和第 4 章中的最优性条件联系起来. 我们的第一个问题是, 对于 (P_ϕ) 形式的凸规划, 找出 Lagrange 乘子和 Lagrange 式的

适当定义．如前面那样，考虑标准凸规划：

(SP) $$\min f(x) \tag{5.109}$$

受限制于

$$g_i(x) \geqslant 0, \quad i=1, \cdots, m, \tag{5.110}$$

并考虑扰动问题的一种可能形式

(PSP) $$\min f(x) \tag{5.111}$$

受限制于

$$g_i(x) \geqslant w_i, \quad i=1, \cdots, m. \tag{5.112}$$

在上两问题中，f 和 g_i 分别是 R^n 上的正常凸函数和正常凹函数．定义

$$\phi_1(x, w)=\begin{cases} f(x), & x \in \mathrm{ED}(f) \text{ 且 } g_i(x) \geqslant w_i,\ i=1, \cdots, m, \\ +\infty, & \text{其他}, \end{cases} \tag{5.113}$$

且定义扰动函数 Φ_1 为

$$\Phi_1(w)=\inf_x \phi_1(x, w). \tag{5.114}$$

现在设 f 是代价函数，代替在无扰动问题(SP)中求 f 的极小，我们可以求解更一般的问题(PSP)，并支付一个量 $\sum_i \lambda_i w_i$，其中 $\lambda_i \geqslant 0$ 是第 i 个约束中每单位扰动所付出的价格．于是我们可以问：在怎样的条件下，无扰动问题(SP)的解依然是最好的选择？回答是，若对一切 $w \in R^m$ 成立

$$\Phi_1(w)-\sum_{i=1}^m \lambda_i w_i \geqslant \Phi_1(0), \tag{5.115}$$

则它依然是最好的选择．这不等式被满足的充要条件是，对所有适合 $g_i(x) \geqslant w_i (i=1, \cdots, m)$ 的 $x \in R^n$ 和 $w \in R^m$，成立

$$f(x)-\sum_{i=1}^m \lambda_i w_i \geqslant \Phi_1(0). \tag{5.116}$$

若我们假定 $\Phi_1(0)$ 有限，则(5.116)等价于条件：对一切 $x \in R^n$，

$$f(x)-\sum_{i=1}^m \lambda_i g_i(x) \geqslant \Phi_1(0). \tag{5.117}$$

(5.117) 的左边是早先几章中所定义的、与 (SP) 相联系的

Lagrange 式，λ_i 是 Lagrange 乘子．使(5.117)左边关于 x 的下确界与(SP)的下确界 $\Phi_1(0)$ 相一致的非负 Lagrange 乘子 λ_i^* 也称为"平衡价格"[31]．(5.117) 的这些解释将证明是与以前几章中导出的鞍点关系相一致的．

在前面讨论的推动下，我们扩充 Lagrange 乘子到更一般的情形．设(P_ϕ)是(5.40)定义的规划，又设 $\Phi(0)$ 有限，则 $\lambda^*\in R^k$ 称为(P_ϕ)的 Lagrange 乘子向量，如果对一切 $w\in R^k$ 有

$$\Phi(w)-\lambda^{*T}w\geqslant\Phi(0). \tag{5.118}$$

类似地，如 $\Psi(0)$ 有限，我们定义 $x^*\in R^n$ 是$(D_{\phi*})$的 Lagrange 乘子，如果对一切 $\xi\in R^n$ 有

$$\Psi(\xi)-\xi^T x^*\geqslant\Psi(0). \tag{5.119}$$

将上面两个不等式与(4.110)比较，可以断定，(5.118)和(5.119)分别等价于条件 $\lambda^*\in\partial\Phi(0)$ 和 $x^*\in\partial\Psi(0)$．

类似地，我们定义相应于(P_ϕ)的 Lagrange 式为

$$L(x,\ \lambda)=\inf_w\{\phi(x,\ w)-\lambda^T w\}. \tag{5.120}$$

这函数也称为相应于(P_ϕ)的"Kuhn-Tucker 函数"，见[19, 31]．在本节的余下部分，我们给出将一对规划(P_ϕ)和$(D_{\phi*})$同 Lagrange 乘子和 Lagrange 式的鞍点联系起来的一系列结果．某些结果将用不同的术语来表达，不可避免地在某些定理中将有所重复．

第一个结果是关于凸规划的 Lagrange 乘子的存在性的．

定理 5.20

假定(P_ϕ)有最优解，则存在 Lagrange 乘子向量 λ^* 的充要条件为(P_ϕ)是稳定的．向量 λ^* 是(P_ϕ)的 Lagrange 乘子向量的充要条件为 $\lambda^*\in\partial\Phi(0)$．

【证明】 如(P_ϕ)有一个最优解，则 $\Phi(0)$ 有限，所以，$\partial\Phi(0)$ 非空的充要条件为(P_ϕ)是稳定的．进一步，从 Lagrange 乘子的定义可推得，λ^* 是 Lagrange 乘子向量的充要条件为 λ^* 是 $\Phi(0)$ 的次梯度． ▌

这个定理有一个有趣的含义．既然可以认为它是对于凸规划

的一般结果，我们断定，对这样的规划，要使 Lagrange 乘子存在，换言之要使 Kuhn-Tucker 条件满足，稳定性是一个充分且必要的品性. 所以，一切不同种类的约束品性在凸规划的情形必定蕴涵着稳定性 上面定理的对偶形式为下面定理.

定理 5.21

假定$(D_{\phi*})$有最优解，则存在 Lagrange 乘子向量 x^* 的充要条件为$(D_{\phi*})$是稳定的. 向量 x^* 是$(D_{\phi*})$的 Lagrange 乘子向量的充要条件为 $x^*\in\partial\Psi(0)$.

现在转向刻划 Lagrange 乘子和最优解的特征，设 ϕ 是 R^{n+k} 上的正常凸函数.

定理 5.22

向量 λ^* 是(P_ϕ)的 Lagrange 乘子向量的充要条件为：λ^* 是$(D_{\phi*})$的最优解且 $\Phi(0)=-\phi^*(0,\lambda^*)$. 对偶地，设 ϕ 是闭的，则 x^* 是$(D_{\phi*})$的 Lagrange 乘子向量的充要条件为：x^* 是(P_ϕ)的最优解且 $\phi(x^*,0)=-\Psi(0)$.

【证明】 由定理 5.20，若 λ^* 是(P_ϕ)的 Lagrange 乘子向量，则 $\lambda^*\in\partial\Phi(0)$. 由定理 5.4，我们有 $0\in\partial\Phi^*(\lambda^*)$. 从次梯度的定义可得，$\Phi^*$ 在 λ^* 达到它的极小值[见(4.110)]. 由引理 5.10，这等价于 λ^* 是$(D_{\phi*})$的最优解. 从练习 4.N，如 $\partial\Phi(0)\neq\emptyset$，则 $\Phi(0)=\mathrm{cl}\,\Phi(0)$. 现在

$$\mathrm{cl}\,\Phi(0)=\Phi^{**}(0)=\sup_\lambda\{0^T\lambda-\Phi^*(\lambda)\}, \tag{5.121}$$

由引理 5.10 我们有

$$\sup_\lambda\{0^T\lambda-\Phi^*(\lambda)\}=\sup_\lambda-\phi^*(0,\lambda)=-\phi^*(0,\lambda^*). \tag{5.122}$$

设 λ^* 是$(D_{\phi*})$的最优解且 $\Phi(0)=-\phi^*(0,\lambda^*)$. 由引理 5.10，这意味着 Φ^* 在 λ^* 达到它的极小值. 所以 $0\in\partial\Phi^*(\lambda^*)$. 由(5.121)和(5.122)，$\Phi(0)=\mathrm{cl}\,\Phi(0)$，由定理 5.4，$\lambda^*\in\partial\Phi(0)$. 这样，$\lambda^*$ 是(P_ϕ)的 Lagrange 乘子向量. 对偶部分的证明是相仿的. ∎

考虑如上定义的 Lagrange 式. 我们将证明这函数的类似于第 3 章和第 4 章的鞍点结果. 首先我们考察 L 的凸凹性质.

引理 5.23

相应于(P_ϕ)的Lagrange式,对任何固定的λ是$x\in R^n$的凸函数,对任何固定的x又是$\lambda\in R^k$的凹函数.

【证明】 对任何固定的λ, L关于x的凸性的证明和定理5.8的证明相同.

设$(x, \lambda^1, \alpha^1)\in Q(L)$, $(x, \lambda^2, \alpha^2)\in Q(L)$,其中$Q(L)$是$L$的下图象,则

$$\alpha^1\leqslant L(x, \lambda^1)=\inf_w\{\phi(x, w)-\lambda^{1T}w\}, \tag{5.123}$$

$$\alpha^2\leqslant L(x, \lambda^2)=\inf_w\{\phi(x, w)-\lambda^{2T}w\}. \tag{5.124}$$

因此

$$q_1\alpha^1+q_2\alpha^2\leqslant\inf_w\{q_1\phi(x, w)-q_1\lambda^{1T}w\}+\inf_w\{q_2\phi(x, w)-q_2\lambda^{2T}w\}, \tag{5.125}$$

也就有

$$q_1\alpha^1+q_2\alpha^2\leqslant\inf_w\{\phi(x, w)-(q_1\lambda^1+q_2\lambda^2)^Tw\}=L(x, q_1\lambda^1+q_2\lambda^2). \tag{5.126}$$

这就是说,对任意固定的x, $(x, q_1\lambda^1+q_2\lambda^2, q_1\alpha^1+q_2\alpha^2)\in Q(L)$,从而$L$关于$\lambda$是凹的. 】

相应于(P_ϕ)的Lagrange式在(5.120)中定义为ϕ的函数,下面我们将看到,可以把这关系倒过来. 下一个引理的证明利用了部分共轭性. 设f是R^{n+k}上的凸函数,对每个固定的$x\in R^n$, $f(x, y)$的部分共轭$f_p^*(x, y)$定义为

$$f_p^*(x, \eta)=\sup_{y\in R^k}\{\eta^Ty-f(x, y)\}, \tag{5.127}$$

而对每个固定的x, $f_p^*(x, \eta)$的部分共轭就是$\mathrm{cl}\,f(x, y)$,其中闭包运算是关于第二个变量的.

引理 5.24

设ϕ是R^{n+k}上的闭正常凸函数,则

$$\phi(x, w)=\sup_\lambda\{L(x, \lambda)+\lambda^Tw\}, \tag{5.128}$$

$$-\phi^*(\xi, \lambda)=\inf_x\{L(x, \lambda)-\xi^T x\}. \tag{5.129}$$

【证明】 我们有

$$\sup_\lambda\{L(x, \lambda)+\lambda^T w\}=\sup_\lambda\{\inf_u[\phi(x, u)-\lambda^T u]+\lambda^T w\} \tag{5.130}$$

$$=\sup_\lambda \{\lambda^T w-\sup_u[\lambda^T u-\phi(x, u)]\} \tag{5.131}$$

$$=\sup_\lambda\{\lambda^T w-\phi_y^*(x, \lambda)\}$$

$$=\phi_y^{**}(x, w)=\phi(x, w). \tag{5.132}$$

类似地,

$$\inf_x \{L(x, \lambda)-\xi^T x\}=\inf_x \inf_u\{\phi(x, u)-\xi^T x-\lambda^T u\} \tag{5.133}$$

$$=-\phi^*(\xi, \lambda). \tag{5.134}$$

】

作为这引理的直接结果，我们得到

$$\phi(x, 0)=\sup_\lambda L(x, \lambda) \tag{5.135}$$

和

$$-\phi^*(0, \lambda)=\inf_x L(x, \lambda). \tag{5.136}$$

在 Lagrange 式的鞍点和相应的凸规划的最优解之间的第一个关系给出在下面定理中.

定理 5.25

设 ϕ 是 R^{n+k} 上的闭正常凸函数, L 是相应于 (P_ϕ) 的 Lagrange 式. 若 (x^*, λ^*) 是 L 的鞍点, 则 x^* 是 (P_ϕ) 的最优解, λ^* 是 (D_{ϕ^*}) 的最优解, 且

$$\phi(x^*, 0)=L(x^*, \lambda^*)=-\phi^*(0, \lambda^*). \tag{5.137}$$

【证明】 若对一切 $x\in R^n$ 和 $\lambda\in R^k$ 有

$$L(x^*, \lambda)\leqslant L(x^*, \lambda^*)\leqslant L(x, \lambda^*), \tag{5.138}$$

则由(5.135)和(5.136)有

$$L(x^*, \lambda^*)=\sup_\lambda L(x^*, \lambda)=\phi(x^*, 0), \tag{5.139}$$

$$L(x^*, \lambda^*)=\inf_x L(x, \lambda^*)=-\phi^*(0, \lambda^*). \tag{5.140}$$

由推论 5.6 可得, x^* 和 λ^* 分别是 (P_ϕ) 和 (D_{ϕ^*}) 的最优解. 】

这结果的逆命题是下列定理.

定理 5.26

设 ϕ 是闭正常凸函数. 若 (P_ϕ) 是稳定的且 x^* 是 (P_ϕ) 的最优解, 则存在 λ^*, 使得 (x^*, λ^*) 是 Lagrange 式 L 的鞍点. 对偶地, 若 (D_{ϕ^*}) 是稳定的且 λ^* 是 (D_{ϕ^*}) 的最优解, 则存在 x^*, 使得 (x^*, λ^*) 是 Lagrange 式 L 的鞍点.

【证明】 若关于 (P_ϕ) 的假定成立, 则由定理 5.11, 存在 (D_{ϕ^*}) 的最优解 λ^*, 且 $\phi(x^*, 0) = -\phi^*(0, \lambda^*)$. 由 (5.135) 和 (5.136) 有

$$\sup_\lambda L(x^*, \lambda) = \inf_x L(x, \lambda^*). \tag{5.141}$$

显然

$$\sup_\lambda L(x^*, \lambda) \geqslant L(x^*, \lambda^*) \geqslant \inf_\lambda L(x, \lambda^*), \tag{5.142}$$

所以对一切 $x \in R^n$ 和 $\lambda \in R^k$, 我们有

$$\begin{aligned} L(x^*, \lambda) &\leqslant \sup_\lambda L(x^*, \lambda) = L(x^*, \lambda^*) \\ &= \inf_x L(x, \lambda^*) \leqslant L(x, \lambda^*). \end{aligned} \tag{5.143}$$

对偶部分的证明是类似的. ▌

下面的定理概括了本节的主要结果, 它的证明是基于前面叙述的结果, 留给读者去完成.

定理 5.27 (等价性定理)

设 ϕ 是 R^{n+k} 上的闭正常凸函数, 则下面五个陈述是等价的:

(i) 向量 (x^*, λ^*) 是相应于 (P_ϕ) 的 Lagrange 式的鞍点.

(ii) 向量 x^* 是 (P_ϕ) 的最优解, 而 λ^* 是 (P_ϕ) 的 Lagrange 乘子向量.

(iii) 向量 λ^* 是 (D_{ϕ^*}) 的最优解, 而 x^* 是 (D_{ϕ^*}) 的 Lagrange 乘子向量.

(iv) 向量 λ^* 和 x^* 分别是 (P_ϕ) 和 (D_{ϕ^*}) 的 Lagrange 乘子向量, 且 $\phi(x^*, 0) = -\phi^*(0, \lambda^*)$.

(v) 规划 (P_ϕ) 和 (D_{ϕ^*}) 是正规的, 且 x^*, λ^* 分别是 (P_ϕ) 和

(D_{ϕ^*})的最优解.

5.4 标准凸规划的对偶性和最优性

现在我们把前几节的结果用于本章之初所定义的标准凸规划,借此来说明这些结果. 为了方便,我们重述标准原有规划如下:

(SP)
$$\min f(x) \tag{5.144}$$
受限制于
$$g_i(x) \geqslant 0, \quad i=1, \cdots, m. \tag{5.145}$$
其中 f 和 g_i 分别是 R^n 上的正常凸函数和凹函数. 注意, 可以通过增加一个"隐式"约束把问题(SP)稍作修改, 例如, 在满足(5.145)的同时, 限制 x 属于一个给定的凸集 $X \subset R^n$. 这一修改使我们可以把(5.145)中一个或几个约束转移到集 X 中, 结果得到对偶变量较少的另一个对偶规划. 一般来说, 这样做没有什么好处, 因为如下面将要看到的, 对偶规划中求极大值的函数这时很少能明显地表示出来. 除非特别提出, 我们总假定 X 就是 R^n.

让我们回到(SP)并引进(5.145)右边的扰动, 扰动问题给出为
$$\min f(x) \tag{5.146}$$
受限制于
$$g_i(x) \geqslant w_i, \quad i=1, \cdots, m, \tag{5.147}$$
扰动函数是
$$\Phi(w)=\inf_x\{f(x): g_i(x) \geqslant w_i,\ i=1, \cdots, m\}. \tag{5.148}$$

为了得到相应于(SP)的对偶规划,我们写下
$$\phi(x, w)=\begin{cases} f(x), & g_i(x) \geqslant w_i,\ i=1, \cdots, m, \\ +\infty, & \text{其他}, \end{cases} \tag{5.149}$$
以后总假定 ϕ 是闭凸函数,且计算它的共轭
$$\phi^*(\xi, \lambda)=\sup_{x, w}\{\xi^T x+\lambda^T w-\phi(x, w)\} \tag{5.150}$$
$$=\sup_{\substack{g_i(x) \geqslant w_i \\ i=1, \cdots, m}}\{\xi^T x+\lambda^T w-f(x)\}. \tag{5.151}$$

引进非负"松弛变量" s_i, 我们可以把约束(5.147)转化为

$$w_i = g_i(x) - s_i, \quad i=1, \cdots, m, \tag{5.152}$$

$$s_i \geqslant 0, \quad i=1, \cdots, m, \tag{5.153}$$

从而得到

$$\phi^*(\xi, \lambda) = \sup_x \left\{ \xi^T x + \sum_{i=1}^m \lambda_i g_i(x) - f(x) \right\} + \sup_{s \geqslant 0} \{-\lambda^T s\}. \tag{5.154}$$

所以

$$\phi^*(\xi, \lambda) = \begin{cases} \sup_x \left\{ \xi^T x + \sum_{i=1}^m \lambda_i g_i(x) - f(x) \right\}, & \lambda \geqslant 0, \\ +\infty, & \text{其他}. \end{cases} \tag{5.155}$$

极大化 $-\phi^*(0, \lambda)$ 所构成的(SP)的对偶问题，它由下式给出：

(DSP) $$\max_{\lambda \geqslant 0} \left\{ \inf_x \left[f(x) - \sum_{i=1}^m \lambda_i g_i(x) \right] \right\}. \tag{5.156}$$

对偶扰动函数 Ψ 是

$$\Psi(\xi) = \inf_{\lambda \geqslant 0} \sup_x \left\{ \xi^T x + \sum_{i=1}^m \lambda_i g_i(x) - f(x) \right\}. \tag{5.157}$$

让我们计算与 (SP) 相联系的 Lagrange 式. 由 (5.120) 和 (5.150), 有

$$L(x, \lambda) = \inf_w \{ f(x) - \lambda^T w : g_i(x) \geqslant w_i, \ i=1, \cdots, m \}, \tag{5.158}$$

象前面一样引进松弛变量 s_i, 我们得到

$$L(x, \lambda) = \begin{cases} f(x) - \sum_{i=1}^m \lambda_i g_i(x), & \lambda \geqslant 0, \\ -\infty, & \text{其他}. \end{cases} \tag{5.159}$$

对于那些使 $L(x, \lambda)$ 取有限值的 x 和 λ, 上面的公式相应于与问题(SP)相联系的 Lagrange 式，如第 2, 3, 4 章所定义的.

(SP) 的 Kuhn-Tucker 型最优性条件现在容易从本章的一般结果导出. 利用稳定性的概念，这些条件变得比前面几章的类似结果更强.

定理 5.28

假定(SP)有最优解 x^*, 则当且仅当(SP)为稳定时，存在向量

$\lambda^* \in R^m$, 使得

$$\lambda^* \geqslant 0, \tag{5.160}$$

$$\lambda_i^* g_i(x^*) = 0, \quad i = 1, \cdots, m, \tag{5.161}$$

且 Lagrange 式 L 有鞍点(x^*, λ^*)，即对所有 $\lambda \geqslant 0$ 和 $x \in R^n$ 成立

$$L(x^*, \lambda) \leqslant L(x^*, \lambda^*) \leqslant L(x, \lambda^*). \tag{5.162}$$

【证明】 设(SP)是稳定的，由定理 5.20，存在 Lagrange 乘子向量 λ^*，使得 $\lambda^* \in \partial\Phi(0)$，即对所有 $w \in R^m$，有

$$\Phi(w) - \lambda^{*T} w \geqslant \Phi(0). \tag{5.163}$$

选取 $w^1 = (-1, 0, \cdots, 0)^T$，则

$$\lambda_1^* \geqslant \Phi(0) - \Phi(w^1). \tag{5.164}$$

因为把 w 的第一个分量从 0 降低到 -1 不能增加 Φ 的值，这不等式的右边是非负的，所以 $\lambda_1^* \geqslant 0$. 令 $w^2 = (0, -1, 0, \cdots, 0)^T$，我们得到 $\lambda_2^* \geqslant 0$，类似地 $\lambda_i^* \geqslant 0$，$i = 1, \cdots, m$. 因为 x^* 对 (SP) 是能行的，我们有

$$\lambda_i^* g_i(x^*) \geqslant 0, \quad i = 1, \cdots, m. \tag{5.165}$$

以 $\widetilde{w}^1 = (g_1(x^*), 0, \cdots, 0)^T$ 代入，就得

$$\Phi(\widetilde{w}^1) - \Phi(0) \geqslant \lambda_1^* g_1(x^*). \tag{5.166}$$

上述不等式的左边必定为 0，这是因为：若 $g_1(x^*) = 0$，则上式中的差恒为 0；若 $g_1(x^*) > 0$，令 w_1 从 0 增加到 $g_1(x^*)$，这并不改变 (SP) 的最优解，因而 Φ 的值不会变得离开 $\Phi(0)$. 这样，$0 \geqslant \lambda_1^* g_1(x^*)$. 类似地有 $0 \geqslant \lambda_i^* g_i(x^*)$，$i = 1, \cdots, m$. 所以(5.161)必定成立. 由定理 5.26，向量(x^*, λ^*)是 L 的鞍点. 反之，设 λ^* 是 Lagrange 乘子向量，若(x^*, λ^*)是 L 对所有 $\lambda \geqslant 0$ 和 $x \in R^n$ 的鞍点，则

$$f(x) - \sum_{i=1}^{m} \lambda_i^* g_i(x) \geqslant f(x^*) - \sum_{i=1}^{m} \lambda_i^* g_i(x^*) = f(x^*), \tag{5.167}$$

其中等式由 (5.161) 推得. 因为 $\lambda^* \geqslant 0$，对于所有满足 $g_i(x) \geqslant w_i$ $(i = 1, \cdots, m)$的 x 和 w，我们得到

$$f(x) - \sum_{i=1}^{m} \lambda_i^* w_i \geqslant f(x^*). \tag{5.168}$$

因为$f(x^*)=\Phi(0)$，取(5.168)左边关于x的下确界，则对所有$w\in \mathrm{ED}(\Phi)$，成立

$$\Phi(w)-\sum_{i=1}^{m}\lambda_i^* w_i \geqslant \Phi(0). \tag{5.169}$$

如果$w \overline{\in} \mathrm{ED}(\Phi)$，我们有$\Phi(w)=+\infty$，从次梯度的定义可知$\lambda^*\in\Phi(0)$. 这样，(SP)是稳定的． ▌

这定理类似于定理4.41，那里我们假定了蕴涵稳定性的强相容性．因为稳定性不蕴涵强相容性，我们不能证明定理4.41的逆命题．但是注意，对于闭的f和g_i，可以对(DSP)问题叙述和证明类似于上面定理的对偶结果．

现在我们转向函数f和g_i可微时的最优性条件．下面的定理构成可微(SP)规划的一组 Kuhn-Tucker 型最优性条件．

定理 5.29

设f和g_1，…，g_m分别是R^n上的实值可微的凸函数和凹函数，假定(SP)有最优解x^*，则当且仅当(SP)为稳定时，存在向量$\lambda^*\in R^m$，使得

$$\nabla f(x^*)-\sum_{i=1}^{m}\lambda_i^*\nabla g_i(x^*)=0, \tag{5.170}$$

$$\lambda_i^* g_i(x^*)=0, \quad i=1, \cdots, m, \tag{5.171}$$

$$\lambda^*\geqslant 0. \tag{5.172}$$

【证明】 设(SP)是稳定的．由前面定理可知(5.171)和(5.172)是成立的，所以我们只需证明(5.170)成立．由定理5.11，(DSP)存在最优解λ^*，且(SP)与(DSP)的最优值是相等的．所以

$$f(x^*)=\inf_x\left\{f(x)-\sum_{i=1}^{m}\lambda_i^* g_i(x)\right\}, \tag{5.173}$$

或者对所有x，有

$$f(x^*)\leqslant f(x)-\sum_{i=1}^{m}\lambda_i^* g_i(x). \tag{5.174}$$

以$x=x^*$代入，可看到(5.174)右边出现的函数在x^*达到它的无约束极小值，所以它的梯度在那里为0，从而(5.170)成立．逆命题部分的证明平行于前面定理的证明，留给读者去完成． ▌

(SP)的对偶规划已由很多学者探讨过，他们提出了若干种描述方式，其中每一种都可以从本章的结果导出．显然，对每个原有凸规划，我们可以表述很多不同的对偶规划，引进一个扰动形式就有一个对偶规划．

对于(SP)，本节所用的扰动形式产生了这样的对偶规划(DSP)，它的目标函数［见(5.156)］是用一个附加的最优化问题来给出的，即表示为

$$\inf_{x}\left\{f(x)-\sum_{i=1}^{m}\lambda_i g_i(x)\right\}. \tag{5.175}$$

但是，对(SP)的某些特殊情形或者扰动的其他形式，这个附加的最优化问题可以明显地解出．这一点将在下面的例子和第7章中说明．

例 5.4.1

让我们用本章的结论导出线性规划的对偶．原有线性规划为

(LP) $$\min c^T x \tag{5.176}$$

受限制于

$$Ax \geqslant b, \tag{5.177}$$

$$x \geqslant 0, \tag{5.178}$$

其中 A 是 $m\times n$ 实矩阵，b 和 c 分别是 m 维和 n 维向量．在约束的右边引进扰动，我们得到

$$\phi(x, w, v)=\begin{cases} c^T x, & Ax-b\geqslant w,\ x\geqslant v, \\ +\infty, & \text{其他}, \end{cases} \tag{5.179}$$

和

$$\phi^*(\xi, \lambda, \nu)=\sup_{\substack{x,w,v\\ Ax-b\geqslant w\\ x\geqslant v}}\{\xi^T x+\lambda^T w+\nu^T v-c^T x\}. \tag{5.180}$$

这表达式简化为

$$\phi^*(\xi, \lambda, \nu)=\begin{cases} \sup\limits_{x}\{(\xi-c+A^T\lambda+\nu)^T x-\lambda^T b\}, & \lambda\geqslant 0,\ \nu\geqslant 0, \\ +\infty, & \text{其他}. \end{cases} \tag{5.181}$$

所以

$$\phi^*(\xi, \lambda, \nu)=\begin{cases} -\lambda^T b, & \lambda \geqslant 0,\ \nu \geqslant 0,\ A^T\lambda+\nu=c-\xi, \\ +\infty, & \text{其他}. \end{cases} \tag{5.182}$$

对偶线性规划成为

(DLP) $$\min \lambda^T b \tag{5.183}$$

受限制于

$$A^T\lambda \leqslant c, \tag{5.184}$$

$$\lambda \geqslant 0, \tag{5.185}$$

其中对偶变量 ν 已经从问题中消去了，因为它只不过是一个松弛变量.

(LP)问题显然是上面(SP)的特殊情况，(DLP)可以从本节的结果得到. 为了这样做，令 $f(x)=c^Tx$, $g(x)=Ax-b$, $X=\{x: x\geqslant 0\}$. 则对偶规划(利用早先讨论过的增加隐式约束来修改(SP)问题)，我们有

$$\max_{\lambda\geqslant 0}\{\inf_{x\in X}[c^Tx-\lambda^T(Ax-b)]\}, \tag{5.186}$$

整理(5.186)，我们得到

$$\max_{\lambda\geqslant 0}\{\inf_{x\geqslant 0}[(c-A^T\lambda)^Tx+\lambda^Tb]\}, \tag{5.187}$$

这样，求极大值的函数当 $c\geqslant A^T\lambda$ 时取值 λ^Tb，而在其他情形取值 $-\infty$.

线性规划的最优性条件和对偶性定理当然也可以从本章的一般结论导出. ∎

我们把本章导出的对偶规划和不用共轭函数所得到的对偶性方面的早期工作加以比较，以结束这一节. 一个典型的例子是下面由 Wolfe[86] 引进的一对对偶问题:

(WP) $$\min_x f(x) \tag{5.188}$$

受限制于

$$g_i(x)\geqslant 0, \quad i=1, \cdots, m, \tag{5.189}$$

和

(WD) $$\max_{x,\lambda}\left\{f(x)-\sum_{i=1}^{m}\lambda_i g_i(x)\right\} \tag{5.190}$$

受限制于

$$\nabla f(x)=\sum_{i=1}^{m}\lambda_i\nabla g_i(x), \quad \lambda\geqslant 0, \tag{5.191}$$

其中 f 和 g_i 分别是 R^n 上的实值可微凸函数和凹函数. 我们立即注意到, Wolfe 的原有问题恰是我们的(SP)形式. 让我们把他的对偶问题(WD)联系到(DSP). 为了方便, 我们重述(SP)的对偶为

$$\text{(DSP)} \qquad \max_{\lambda\geqslant 0}\left\{\inf_x\left[f(x)-\sum_{i=1}^{m}\lambda_i g_i(x)\right]\right\}. \tag{5.192}$$

首先我们注意, 给定任意 $\hat{\lambda}\geqslant 0$, 向量 $\hat{x}$ 对 Wolfe 的对偶问题是能行的充要条件为, $\hat{x}$ 是 $f(x)-\sum_i \hat{\lambda}_i g_i(x)$ 的整体极小值点. 这条件是从 f 的凸性, g_i 的凹性和 (5.191) 的梯度条件推得的. 这样, (5.190)和(5.191)可以等价地重新写成问题

$$\text{(WSD)} \qquad \max_{\lambda\geqslant 0}\left\{\min_x\left[f(x)-\sum_{i=1}^{m}\lambda_i g_i(x)\right]\right\}, \tag{5.193}$$

其中 λ 进一步限制为使上式中的极小在某个 x 达到的那些值.

我们可以看到, 如果 λ 对(WSD)能行, 则它对(DSP)也能行, 且求极大值的函数是相等的. 但是如果最优解 λ^* 对 (WSD) 不是能行的, 那么 (DSP) 的最优值可能大于 (WSD) 的最优值. 在 Wolfe 所研究的条件下, 这种情形不会出现. 我们首先建立下面的引理, 它从以前的结果直接可推得.

引理 5.30

设(WP)是稳定的, x^* 是一个最优解, 则 x^* 是

$$f(x)-\sum_{i=1}^{m}\lambda_i^* g_i(x)$$

的整体极小值点, 其中 λ^* 是(DSP)的最优解.

【证明】 设 x^* 是(WP)的最优解, 由定理 5.11, (DSP)有最优解 λ^*, 从而有

$$f(x^*)=\inf_x\left\{f(x)-\sum_{i=1}^{m}\lambda_i^* g_i(x)\right\}. \tag{5.194}$$

由定理 5.22 和 5.28 可知, 对 $i=1, \cdots, m$ 有 $\lambda_i^* g^i(x^*)=0$, 因此

$$f(x^*)-\sum_{i=1}^{m}\lambda_i^* g_i(x^*)=\inf_x\left\{f(x)-\sum_{i=1}^{m}\lambda_i^* g_i(x)\right\}, \quad (5.195)$$

且我们可以用极小值代替下确界. ∎

从这引理我们断定，若(WP)稳定且有最优解，则(WSP)和(DSP)的最优解是相同的. Wolfe 在 Kuhn-Tucker 或某种其他的约束品性[22,25]对(WP)成立的条件下，叙述他的强对偶性定理，这条件蕴涵着(WP)是稳定的[18].

定理 5.31(Wolfe)

如果 x^* 为(WP)的解，那么存在 λ^*，使得(x^*, λ^*)为(WD)的解，且它们的最优值相等.

上述问题(WD)除了特殊情形外，在计算上不是很吸引人的. 虽然(WP)是凸规划，(WD)一般是非凸的. 它含有比原有规划更多的变量，它的目标函数关于(x, λ)不是凹的，它的能行集不一定是凸的，与 Wolfe 的对偶性关系相类似的讨论可以在 Geoffrion[18]和 Whinston[34] 中找到.

如果 (WP) 中的约束 (5.189) 全是线性的，我们就得到 Dennis[8] 和 Dorn[10] 研究过的一对对偶问题. 前一工作是有趣的，因为它利用 Legendre 变换，我们已看到 Legendre 变换与共轭函数是密切相关的. Wolfe 关于非线性凸规划的对偶性的处理，是这领域中在 Rockafellar 的结果之前一些工作的代表性例子. 所有这些早期工作强烈地依靠非线性规划中具有或不具有可微性的 Kuhn-Tucker 理论[22]，这理论本身是经典最优化的 Lagrange 乘子理论的扩充. 在这些著作中，我们可以举出 Cottle[3]，Dantzig、Eisenberg、Cottle[6]，Eisenberg[11]，Hanson[20]，Huard[21]，Mangasarian[24]，Mangasarian、Ponstein[26] 和 Stoer[33] 的工作. 有趣的是，和早期工作不同，我们先建立对偶性关系，然后从中获得 Kuhn-Tucker 型的最优性条件.

线性规划中对偶性的经济解释已被推广到非线性凸规划. 有兴趣的读者可以阅读Balinski、Baumol[1]，Gale[14,15]和Williams[35]在这方面的著作. Williams 的工作是基于共轭函数，且对本章导

出的很多结果给出经济意义.

练　习

5.A. 设 f 是 R^n 上的具有实值或无穷值的任意函数，证明它的共轭函数 f^* 是凸的.

5.B. 给定 R^n 上的两个正常凸函数 f_1, f_2，定义它们的结合函数为

$$(f_1 \square f_2)(x)=\inf_y\{f_1(x-y)+f_2(y)\}, \tag{5.196}$$

证明 $f_1 \square f_2$ 是凸的，且证明它的共轭是

$$(f_1 \square f_2)^*(\xi)=f_1^*(\xi)+f_2^*(\xi), \tag{5.197}$$

找出结合函数及其共轭的经济解释.

5.C. 给定正齐次凸函数，证明它的共轭函数是一个指示函数. 试找出一个指示函数，使得它的共轭也是一个指示函数.

5.D. 函数 f 由下式给出：

$$f(x)=\begin{cases}\left(\sum_{j=1}^n (x_j)^p\right)^{1/p}, & x\geqslant 0,\\ +\infty, & \text{其他},\end{cases} \tag{5.198}$$

其中 p 是非零实常数，证明 f 或者是凸的，或者是凹的，这取决于 p 的值，并求 f 的共轭. 试求当 $p\to\pm\infty$ 时 f 和 f^* 的极限情形.

5.E. R^n 上的函数 f 称为可分离的，如果

$$f(x)=f_1(x_1)+f_2(x_2)+\cdots+f_n(x_n), \tag{5.199}$$

证明它的共轭也是可分离的.

5.F. 设 A 是 n 阶非异实阵，b 和 c 是 R^n 中的向量，α 是实数，定义 $\tilde{f}$ 为

$$\tilde{f}(x)=f(Ax-b)+c^Tx+\alpha, \tag{5.200}$$

其中 f 是凸函数. 求 $\tilde{f}^*$.

5.G. 证明：如果凸规划 (P_ϕ) 的扰动函数 Φ 是正常的，且 $0\in \mathrm{ri}[\mathrm{ED}(\Phi)]$，则规划 (P_ϕ) 是稳定的.

5.H. 证明定理 5.16.

5.I. 讨论例 5.2.2 中规划 $(D_{\phi*})$ 的稳定性.

5.J. 令

$$\phi(x, w)=\begin{cases}e^{-\sqrt{xw}}, & x\in R,\ x\geqslant 0,\ w\in R,\ w\geqslant 0,\\ +\infty, & \text{其他},\end{cases} \tag{5.201}$$

讨论 (P_ϕ) 及其对偶的性质.

5.K. 为什么对定理 5.22 的第二部分我们要求 ϕ 是闭的? 完成其证明.

5.L. 证明定理 5.27.

5.M. 完成定理 5.29 的证明。

5.N. 给定原有凸规划

$$\min\{f(x)-g(x)\}, \tag{5.202}$$

其中 f 和 g 分别是 R^n 上的正常凸函数和凹函数，以 $f(x+w)$ 代替 $f(x)$ 产生扰动，试求相应的对偶规划。

5.O. 给定原有凸规划

$$\min_{x\in K} f(x), \tag{5.203}$$

其中 f 是闭正常凸函数，$K\subset R^n$ 是非空闭凸锥，证明对偶规划可表示为

$$\max_{\lambda\in K^*}\{-f^*(\lambda)\}, \tag{5.204}$$

其中 K^* 是 K 的正法锥(见第 3 章)，即

$$K^*=\{\lambda:\lambda\in R^n,\ \lambda^T x\geqslant 0,\ \forall\, x\in K\}. \tag{5.205}$$

参考文献

1. Balinski, M. L., and W. J. Baumol, "The Dual in Nonlinear Programming and Its Economic Interpretation," *Review of Economic Studies*, 35, 237–256 (1968).
2. Bartle, R. G., *The Elements of Real Analysis*, 2nd ed., John Wiley & Sons, New York, 1976.
3. Cottle, R. W., "Symmetric Dual Quadratic Programs," *Quart. Appl. Math.*, 21, 237–243 (1963).
4. Courant, R., and D. Hilbert, *Methods of Mathematical Physics*, Interscience Publishers, New York, Vol. I (1953), Vol. II (1962).
5. Dantzig, G. B., *Linear Programming and Extensions*, Princeton University Press, Princeton, N.J., 1963.
6. Dantzig, G. B., E. Eisenberg, and R. W. Cottle, "Symmetric Dual Nonlinear Programs," *Pacific J. of Math.*, 15, 809–812 (1965).
7. Dantzig, G. B., J. Folkman, and N. Z. Shapiro, "On the Continuity of the Minimum Set of a Continuous Function," *J. Math. Anal. & Appl.*, 17, 519–548 (1967).
8. Dennis, J. B., *Mathematical Programming and Electrical Networks*, John Wiley & Sons, New York, 1959.
9. Dorfman, R., P. Samuelson, and R. Solow, *Linear Programming and Economic Analysis*, McGraw-Hill Book Co., New York, 1958.
10. Dorn, W. S., "A Duality Theorem for Convex Programs," *IBM J. Res. & Dev.*, 4, 407–413 (1960).

11. EISENBERG, E., "Duality in Homogeneous Programming," *Proc. Am. Math. Soc.*, 12, 783–787 (1961).

12. EVANS, J. P., and F. J. GOULD, "Stability in Nonlinear Programming," *Operations Research*, 18, 107–118 (1970).

13. FENCHEL, W., "On Conjugate Convex Functions," *Can. J. Math.*, 1, 73–77 (1949).

14. GALE, D., "A Geometric Duality Theorem with Economic Applications," *Review of Economic Studies*, 34, 19–24 (1967).

15. GALE, D., "Non-Linear Duality and Qualitative Properties of Optimal Growth," in *Integer and Nonlinear Programming*, J. Abadie (Ed.), North-Holland Publishing Co., Amsterdam, 1970.

16. GALE, D., H. W. KUHN, and A. W. TUCKER, "Linear Programming and the Theory of Games," in *Activity Analysis of Production and Allocation*, T. C. Koopmans (Ed.), John Wiley & Sons, New York, 1951.

17. GELFAND, I. M., and S. V. FOMIN, *Calculus of Variations*, Prentice-Hall, Englewood Cliffs, N.J., 1963.

18. GEOFFRION, A. M., "Duality in Nonlinear Programming: A Simplified Applications-Oriented Development," *SIAM Review*, 13, 1–37 (1971).

19. HAMALA, M., "Geometric Programming in Terms of Conjugate Functions," Center for Operations Research and Econometrics, Université Catholique de Louvain Discussion Paper No. 6811, Louvain, Belgium, June 1968.

20. HANSON, M. A., "A Duality Theorem in Non-Linear Programming with Non-Linear Constraints," *Australian J. Statistics*, 3, 64–72 (1961).

21. HUARD, P., "Dual Programs," *IBM J. Res. & Dev.*, 6, 137–139 (1962).

22. KUHN, H. W., and A. W. TUCKER, "Nonlinear Programming," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman (Ed.), University of California Press, Berkeley, Calif., 1951.

23. KUHN, H. W., and A. W. TUCKER (Eds.), *Linear Inequalities and Related Systems*, Princeton University Press, Princeton, N.J., 1956.

24. MANGASARIAN, O. L., "Duality in Nonlinear Programming," *Quart. Appl. Math.*, 20, 300–302 (1962).

25. MANGASARIAN, O. L., *Nonlinear Programming*, McGraw-Hill Book Co., New York, 1969.

26. MANGASARIAN, O. L., and J. PONSTEIN, "Minmax and Duality in Nonlinear Programming," *J. Math. Anal. & Appl.*, 11, 504–518 (1965).

27. RISSANEN, J., "On Duality without Convexity," *J. Math. Anal. & Appl.*, 18, 269–275 (1967).

28. RISSANEN, J., "Reproof of a Theorem on Duality without Convexity," *J. Math. Anal. & Appl.*, 29, 429–431 (1970).

29. Rockafellar, R. T., "Duality and Stability in Extremum Problems Involving Convex Functions," *Pacific J. Math.*, **21**, 167–187 (1967).

30. Rockafellar, R. T., "Conjugates and Legendre Transforms of Convex Functions," *Can. J. Math.*, **19**, 200–205 (1967).

31. Rockafellar, R. T., "Duality in Nonlinear Programming," in *Mathematics of the Decision Sciences*, Part 1, G. B. Dantzig and A. F. Veinott (Eds.), American Mathematical Society, Providence, R. I., 1968.

32. Rockafellar, R. T., *Convex Analysis*, Princeton University Press, Princeton, N.J., 1970.

33. Stoer, J., "Duality in Nonlinear Programming and the Minimax Theorem," *Numerische Mathematik*, **5**, 371–379 (1963).

34. Whinston, A., "Some Applications of the Conjugate Function Theory to Duality," in *Nonlinear Programming*, J. Abadie (Ed.), North-Holland Publishing Co., Amsterdam, 1967.

35. Williams, A. C., "Nonlinear Activity Analysis and Duality," in *Proceedings of the Princeton Symposium on Mathematical Programming*, H. W. Kuhn (Ed.), Princeton University Press, Princeton, N.J., 1970.

36. Wolfe, P., "A Duality Theorem for Non-Linear Programming," *Quart. Appl. Math.*, **19**, 239–244 (1961).

第 6 章

广 义 凸 性

在前两章中，我们从数学规划的观点讨论了凸性的一些最重要的方面．对那里所介绍的结果来说，凸性是一个充分条件．但是很清楚，它决不是一个必要条件．在本章中，我们要扩充凸性的概念，从而使先前所得的大多数结果能容易地被推广．

本章讨论的主题仍在发展，实际上，这个方向上的研究已构成数学规划的当前趋向之一．幸而，研究凸集和凸函数时所引进的工具，在把前两章的结果推广到某些类型的非凸规划上去时，一般说是足够的．这样的推广是很直接的，因为我们从凸分析中已了解到有希望获得哪种结果．因此，我们仅介绍关于凸性各种推广的基本思想，有兴趣的读者能据此导出更完整的理论．下一章，我们将利用本章和前几章的结果，分析一些特殊的非线性规划．

除了广义凸分析还需要研究之外，更需要在寻找非凸最优化问题的有效数值解法方面进行研究，甚至在非凸算法领域中，广义凸性的影响迄今尚未感觉到．我们希望，本章介绍的结果将促进未来的工作，以发展新的数值方法，并扩充关于广义凸最优化问题的现有方法的收敛性结果．

6.1 拟凸函数和伪凸函数

给定 $X\subset R^n$ 上的一个实值函数 f 与数 $\alpha\in R$，定义 f 的水平集 $S(f, \alpha)$ 为

$$S(f, \alpha)=\{x: x\in X, f(x)\leqslant\alpha\}. \tag{6.1}$$

容易证明，若 X 是凸集而 f 是凸函数，则对每个 $\alpha\in R$，$S(f, \alpha)$ 都是凸集．反之，假设现在对于某个 f，已知 $S(f, \alpha)$ 对每个 $\alpha\in R$ 都是凸的，由此是否可推出 f 是凸的？读者能立即否定这个断言．简单的例子是 $X=\{x: x\in R, x\geqslant0\}$ 和凹函数 $f(x)=(x)^{1/2}$．即使

如此,我们仍能用即将给出的凸性的第一种扩充,来刻划水平集均为凸集的函数的特征.

定义在凸集 $X \subset R^n$ 上的实值函数 f 称为是拟凸的, 如果对任意 $x^1 \in X$, $x^2 \in X$ 和权 q (回忆一下,第4章中权的性质)有

$$f(q_1 x^1 + q_2 x^2) \leqslant \max[f(x^1), f(x^2)]. \tag{6.2}$$

显然,每个实值的凸函数都是拟凸的,但反之不然. 函数 f 称为是拟凹的, 如果 $-f$ 是拟凸的. 关于拟凸函数的第一个结果是下列定理.

定理 6.1

设 f 是定义在凸集 $X \subset R^n$ 上的实值函数,则对每个 $\alpha \in R$, f 的水平集都是凸集的充分必要条件为 f 是拟凸函数.

【证明】 设对每个 $\alpha \in R$, $S(f, \alpha)$ 都是凸集. 令 $x^1 \in X$, $x^2 \in X$, $\bar{\alpha} = \max[f(x^1), f(x^2)]$, 则 $x^1 \in S(f, \bar{\alpha})$, $x^2 \in S(f, \bar{\alpha})$. 因为 $S(f, \bar{\alpha})$ 是凸的, 所以对任意权 q 也有 $(q_1 x^1 + q_2 x^2) \in S(f, \bar{\alpha})$.

因此,

$$f(q_1 x^1 + q_2 x^2) \leqslant \bar{\alpha} = \max[f(x^1), f(x^2)]. \tag{6.3}$$

反之, 令 $S(f, \alpha)$ 是 f 的任意水平集. 若 $x^1 \in S(f, \alpha)$, $x^2 \in S(f, \alpha)$, 则

$$f(x^1) \leqslant \alpha, \quad f(x^2) \leqslant \alpha. \tag{6.4}$$

因为 f 是拟凸的,我们有

$$f(q_1 x^1 + q_2 x^2) \leqslant \alpha, \tag{6.5}$$

因而 $(q_1 x^1 + q_2 x^2) \in S(f, \alpha)$. ▌

具有凸水平集的函数曾经由 de Finetti[15] 研究过, 他的结果后来被 Fenchel[17] 推广和校正.

图 6.1 说明了凸函数和拟凸函数之间的一些主要差别. 与凸函数相反,拟凸函数在它的定义域内部可以有不连续性,而且并非每个局部极小必是一个整体极小. 然而, 非整体的局部极小不能是严格极小,这一点我们能在下述结果[24]中看到.

定理 6.2

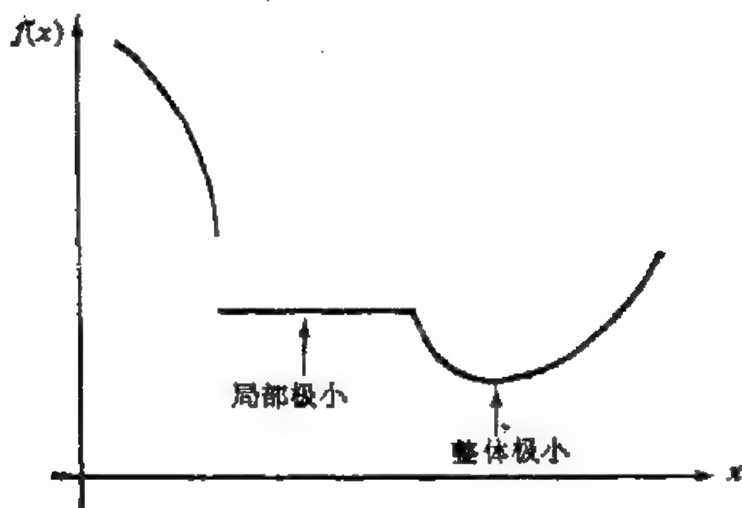


图 6.1 一个拟凸函数

设 f 是在凸集 $X\subset R^n$ 上的拟凸函数．若 x^* 是 f 的一个严格局部极小值点，则 x^* 也是 f 在 X 上的严格整体极小值点．

【证明】 同前面一样，记

$$N_\delta(x^*)=\{x:x\in R^n,\ \|x-x^*\|<\delta\}. \qquad (6.6)$$

设 x^* 是一个严格局部极小值点——即存在 $\delta>0$，使得对集合 $X\cap N_\delta(x^*)$ 中的每个 $x\neq x^*$ 有

$$f(x)>f(x^*). \qquad (6.7)$$

现在设 x^* 不是 f 在 X 上的严格整体极小值点——即存在 $\bar{x}\in X$，$\bar{x}\neq x^*$，使得

$$f(\bar{x})\leqslant f(x^*). \qquad (6.8)$$

于是由拟凸性，对所有权 q 有

$$f(q_1\bar{x}+q_2x^*)\leqslant f(x^*). \qquad (6.9)$$

但对于充分小的 q_1，$(q_1\bar{x}+q_2x^*)\in X\cap N_\delta(x^*)$，这就与(6.7)相矛盾． 】

若拟凸函数是可微的，则能导出进一步的特性．

定理 6.3

设 f 在开凸集 $X\subset R^n$ 上是可微的，则 f 是拟凸函数的充要条件为：对任意两点 $x^1\in X$，$x^2\in X$，如满足

$$f(x^1)\leqslant f(x^2), \qquad (6.10)$$

就必有

$$(x^1-x^2)^T\nabla f(x^2)\leqslant 0. \tag{6.11}$$

【证明】 设 f 是拟凸的，$x^1\in X$，$x^2\in X$，则 $f(x^1)\leqslant f(x^2)$ 蕴涵：对任意 λ，$1\geqslant\lambda\geqslant 0$ 有

$$f(\lambda x^1+(1-\lambda)x^2)\leqslant f(x^2). \tag{6.12}$$

因为 f 是可微的，得

$$Df(x^2;\ x^1-x^2)=(x^1-x^2)^T\nabla f(x^2), \tag{6.13}$$

又有

$$Df(x^2;\ x^1-x^2)=\lim_{\lambda\to 0^+}\frac{f(x^2+\lambda(x^1-x^2))-f(x^2)}{\lambda}. \tag{6.14}$$

由(6.12)至(6.14)，可得

$$Df(x^2;\ x^1-x^2)=(x^1-x^2)^T\nabla f(x^2)\leqslant 0. \tag{6.15}$$

反之，设 $f(x^1)\leqslant f(x^2)$，又设存在一个 $x^3=\lambda x^1+(1-\lambda)x^2$ $(0<\lambda<1)$，使得

$$f(x^2)<f(x^3), \tag{6.16}$$

则根据假设，有

$$(x^2-x^3)^T\nabla f(x^3)\leqslant 0, \tag{6.17}$$

$$(x^1-x^3)^T\nabla f(x^3)\leqslant 0. \tag{6.18}$$

把 x^3 写成 x^1 和 x^2 的凸组合，代入(6.17)和(6.18)，整理后得

$$(x^2-x^1)^T\nabla f(x^3)\leqslant 0, \tag{6.19}$$

$$(x^1-x^2)^T\nabla f(x^3)\leqslant 0. \tag{6.20}$$

因此

$$(x^2-x^1)^T\nabla f(x^3)=0. \tag{6.21}$$

定义

$$U=\{x:f(x)\leqslant f(x^2),\ x=\mu x^2+(1-\mu)x^3,\ 1\geqslant\mu\geqslant 0\}. \tag{6.22}$$

于是我们能找到最接近于 x^3 的一个 $x^0\in U$. 又根据中值定理[8]，能找到 x^0 和 x^3 之间的一个 $\tilde{x}$，使得对于某个 μ^0，$1\geqslant\mu^0\geqslant 0$，有

$$f(x^3)=f(x^0)+(x^3-x^0)^T\nabla f(\tilde{x}) \tag{6.23}$$

$$=f(x^0)+\mu^0(x^3-x^2)^T\nabla f(\tilde{x}), \tag{6.24}$$

因而

$$f(x^3)=f(x^0)+\lambda\mu^0(x^1-x^2)^T\nabla f(\tilde{x}). \tag{6.25}$$

因为 $\tilde{x}$ 也是 x^1 和 x^2 的凸组合，且 $f(x^2)<f(\tilde{x})$，所以由(6.16)至(6.21)，可推出

$$(x^1-x^2)^T\nabla f(\tilde{x})=0. \tag{6.26}$$

因此 $f(x^3)=f(x^0)\leqslant f(x^2)$，这就与(6.16)相矛盾. ▌

对拟凹情况，不等式(6.10)和(6.11)中的不等号要反向.

在二次可微函数的情形，Arrow、Enthoven[2] 导出了用函数的加边行列式表示的在 R^n 的非负象限上拟凸性的充分必要条件. 这些条件后来由 Ferland[18] 扩充到更一般凸集上的函数. 一个二次可微函数 f 在一点 $x\in R^n$ 处的 k 阶加边行列式 $D_k(f, x)$ 定义为

$$D_k(f, x)=\det\begin{bmatrix} 0 & \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_k} \\ \frac{\partial f}{\partial x_1} & \frac{\partial^2 f}{\partial x_1\partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1\partial x_k} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f}{\partial x_k} & \frac{\partial^2 f}{\partial x_k\partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_k\partial x_k} \end{bmatrix}, \quad k=1, \cdots, n. \tag{6.27}$$

如果集合 $X\subset R^n$ 具有非空的内部，称它为实心的(R^n 的非负象限显然是实心集). 函数 f 在实心凸集 $X\subset R^n$ 上是拟凸的必要条件为：对每个 $x\in X$ 有

$$D_k(f, x)\leqslant 0, \quad k=1, \cdots, n. \tag{6.28}$$

类似地，若 f 在实心凸集 $X\subset R^n$ 上是拟凹的，则对每个 $x\in X$ 有

$$(-1)^k D_k(f, x)\geqslant 0, \quad k=1, \cdots, n. \tag{6.29}$$

反之，考虑在包含凸集 $X\subset R^n$ 的一个开集上的二次连续可微的实值函数 f. 若对每个 $x\in X$ 有

$$D_k(f, x)<0, \quad k=1, \cdots, n, \tag{6.30}$$

则 f 在 X 上是拟凸的；若对每个 $x\in X$ 有

$$(-1)^k D_k(f, x)>0, \quad k=1, \cdots, n, \tag{6.31}$$

则 f 在 X 上是拟凹的.

让我们增强拟凸性的要求. 定义在凸集 $X\subset R^n$ 上的实值函

数 f 称为是强拟凸的，如果对任意 $x^1\in X$, $x^2\in X$, $x^1\neq x^2$, 和任意权 $q_1>0$, $q_2>0$, $q_1+q_2=1$, 有

$$f(q_1x^1+q_2x^2)<\max[f(x^1), f(x^2)]. \tag{6.32}$$

对于这种函数,由上述定义立即推出下述结果.

定理 6.4

定义在凸集 $X\subset R^n$ 上的强拟凸函数 f, 至多在一点达到它在 X 上的极小值.

读者能容易地验证每个严格凸函数也是强拟凸的.

我们能稍减轻强拟凸性的概念. 定义在凸集 $X\subset R^n$ 上的实值函数 f 称为是**严格拟凸的**,当且仅当:对任意 $x^1\in X$, $x^2\in X$ 且 $f(x^1)\neq f(x^2)$ 的两点和对任意 $q_1>0$, $q_2>0$, $q_1+q_2=1$ 成立(6.32)式[20]. 容易证明,凸函数和强拟凸函数也是严格拟凸的. 虽然强拟凸函数也是拟凸的,但严格拟凸函数不一定是拟凸的,因为函数

$$f(x)=\begin{cases}1, & x=0,\\ 0, & x\neq 0\end{cases} \tag{6.33}$$

在 $X=R$ 上是严格拟凸,但它不是拟凸的. 然而已经证明,当函数是下半连续时,由严格拟凸性可推出拟凸性[23]. 下列定理给出了严格拟凸函数的很好的性质.

定理 6.5

设 f 是在凸集 $X\subset R^n$ 上的严格拟凸函数, $x^*\in X$ 是 f 的一个局部极小值点,则 x^* 也是 f 在 X 上的整体极小值点.

证明十分类似于定理 6.2, 从略. 在这一章的后面部分将用这个定理中所叙述的性质来刻划函数.

在第 3 章中我们知道,对一般非线性规划,若无某些正则性条件,则最优解的 Kuhn-Tucker 必要条件可以不成立. 在凸规划情形, 第 4 章所述的强相容性的 Slater 条件保证了这些必要条件在最优点确实成立. Arrow、Enthoven[2] 和 Arrow、Hurwicz、Uzawa[4] 已经把 Slater 条件以及定理 4.39 扩充到更一般的规划. 设有一个在 3.1 节中定义的非线性规划(P), 它的约束为

$$g_i(x) \geqslant 0, \quad i=1, \cdots, m, \tag{6.34}$$

$$h_j(x)=0, \quad j=1, \cdots, p, \tag{6.35}$$

其中 g_i 和 h_j 分别为可微的拟凹函数和线性函数. 如果这个规划是强相容的——即存在一个 $x^0 \in R^n$, 它满足(6.34)和(6.35)两式,使得不等式都是严格不等的, 且 h_j 的系数向量是线性无关的, 又

$$\nabla g_i(x^*) \neq 0, \quad i \in I(x^*)=\{i: g_i(x^*)=0\}, \tag{6.36}$$

那末原来对凸规划导出的定理4.39, 在这个更一般的情况下也成立.

这里要注意,满足(6.34)和(6.35)的 $x \in R^n$ 之集合为一凸集. 因此,如果求凸函数在这些约束条件下的极小,则每个局部极小也是整体极小. 这个结果以后将推广到目标函数为其他类型的情形. 还要注意, 具有拟凹约束的一般非线性规划可以具有这样的点,它不是整体解,但满足最优性必要条件. 关于凸函数与拟凸函数性质的有趣的比较,读者可参看 Greenberg、Pierskalla[19].

现在我们转向凸性的另一种扩充, 它是由 Mangasarian[25,26] 引入的.

可微的实值函数 f 称为在开凸集 $X \subset R^n$ 上是伪凸的,如果对任两点 $x^1 \in X$、$x^2 \in X$,有下列性质:

$$(x^1-x^2)^T \nabla f(x^2) \geqslant 0 \text{ 蕴涵 } f(x^1) \geqslant f(x^2). \tag{6.37}$$

函数 f 称为是伪凹的, 如果 $-f$ 是伪凸的, 或等价地,(6.37)中的两个不等式均反向. 如果 f 既是伪凸又是伪凹的, 有时称它为伪线性的. 有趣的是含有线性函数的数学规划的某些性质能推广到含有伪线性函数的规划上去[23]. 可微的凸函数是伪凸的, 而伪凸函数全是严格拟凸的. 因此, 伪凸函数的每个局部极小也即是整体极小. 虽然严格拟凸函数可以有逗留点,在那里梯度为0, 却并不是整体极小值点,但这种情况对伪凸函数却是不可能的.

定理 6.6

设 f 是开凸集 $X \subset R^n$ 上的伪凸函数, 又设对某个 $x^* \in X$, $\nabla f(x^*)=0$,则 x^* 是 f 在 X 上的整体极小值点.

【证明】 由伪凸函数的定义可立即推得. ∎

Ferland[15] 证明了，对二次连续可微函数，前面讲到的以行列式形式给出的拟凸性的充分条件，也是伪凸性的充分条件. 他证明了，如果 f 在实心凸集 $X\subset R^n$ 上是二次可微的拟凸函数，则 f 在任何 $x^0\in X$，$\nabla f(x^0)\neq 0$ 处是伪凸的. 即对每个 $x\in X$，$(x-x^0)^T\nabla f(x^0)\geqslant 0$ 蕴涵 $f(x)\geqslant f(x^0)$.

我们能对伪凸函数作进一步的限制，要求(6.37)中被推出的不等式对 $x^1\neq x^2$ 是严格不等的. 在这种情况下，该函数称为**严格伪凸函数**. 而定理 6.6 就能更强地叙述成：如果 $\nabla f(x^*)=0$，则 x^* 是 f 在 X 上的唯一整体极小值点. 要想考察至今所定义的各类广义凸函数之间的关系，读者可参看 Ponstein[37]，那里还研究了更多的广义凸函数类.

我们现要证明，对于定理 3.8 所述的非线性规划中最优性的 Kuhn-Tucker 必要条件，当目标函数和约束函数都是某种类型的广义凸函数或广义凹函数时，它也是充分条件. 下面结果则是定理 4.38 之推广. 考察非线性规划：

(PQP) $$\min f(x) \tag{6.38}$$

受限制于约束

$$g_i(x)\geqslant 0,\quad i=1,\ \cdots,\ m, \tag{6.39}$$

$$h_j(x)=0,\quad j=1,\ \cdots,\ p. \tag{6.40}$$

象以前一样，令

$$I(x^0)=\{i: g_i(x^0)=0,\quad i=1,\ \cdots,\ m\}. \tag{6.41}$$

于是我们有下述结果[35]. 它的另一种稍加限制的形式可在 [2] 中找到，那里还讨论了拟凸性的经济学方面的内容.

定理 6.7

设 f, g_1, $\cdots$, g_m 和 h_1, $\cdots$, h_p 是在某开凸集 $X\subset R^n$ 上的可微实值函数. 假设 f 是伪凸的，g_i, $i\in I(x^*)$ 是拟凹的，而 h_j 既是拟凸又是拟凹的. 如果存在向量 $\lambda^*\in R^m$、$\mu^*\in R^p$，使得下列诸式成立：

$$\nabla f(x^*)-\sum_{i=1}^{m}\lambda_i^*\nabla g_i(x^*)-\sum_{j=1}^{p}\mu_j^* h_j(x^*)=0, \tag{6.42}$$

$$\lambda_i^* g_i(x^*)=0,\quad i=1,\cdots,m, \tag{6.43}$$

$$g_i(x^*)\geqslant 0,\quad i=1,\cdots,m, \tag{6.44}$$

$$h_j(x^*)=0,\quad j=1,\cdots,p, \tag{6.45}$$

$$\lambda^*\geqslant 0, \tag{6.46}$$

那末 x^* 是(PQP)的整体最优解.

【证明】 设 $x\in X$ 是满足(6.39)和(6.40)的任一点,则

$$g_i(x)\geqslant g_i(x^*),\quad i\in I(x^*), \tag{6.47}$$

$$h_j(x)=h_j(x^*),\quad j=1,\cdots,p. \tag{6.48}$$

因此有

$$(x-x^*)^T\nabla g_i(x^*)\geqslant 0,\quad i\in I(x^*), \tag{6.49}$$

$$(x-x^*)^T\nabla h_j(x^*)=0,\quad j=1,\cdots,p. \tag{6.50}$$

注意到对于 $i\bar{\in} I(x^*)$ 有 $\lambda_i^*=0$,由此得出

$$(x-x^*)^T\left\{\sum_{i=1}^{m}\lambda_i^*\nabla g_i(x^*)+\sum_{j=1}^{p}\mu_j^*\nabla h_j(x^*)\right\}\geqslant 0, \tag{6.51}$$

又由(6.42)便得

$$(x-x^*)^T\nabla f(x^*)\geqslant 0. \tag{6.52}$$

因为 f 是伪凸的,由(6.37)得 $f(x)\geqslant f(x^*)$. ∎

Minch[34] 定义了实值函数 f 的对称梯度 $f^s(x)$,用它来扩充可微的凸、伪凸和拟凸函数的概念. 在一个变量函数的情况下,$f^s(x)$由下式给出:

$$f^s(x)=\lim_{h\to 0}\frac{f(x+h)-f(x-h)}{2h}, \tag{6.53}$$

这里假定上述极限存在. 例如,函数 $f(x)=|x|$ 在 $x=0$ 处是不可微的,但是它在那里具有对称梯度 $f^s(0)=0$. 对于凸函数和至今所讨论过的它的扩充,把通常的梯度换成对称梯度,Minch 已得到了类似于可微情形的结果.

附带提一下,凸函数在对策论的所谓极小极大定理中也起重要的作用,这些定理中的第一个在 1928 年由 von Neumann 叙述. 后来这些定理被推广,证明了对于拟凸函数情形也成立,参看

Berge[12], Nikaidô[35].

尽管前面所讲的一些结果有明显的作用，但是确立函数的拟凸性和伪凸性的困难问题依然存在。这种困难在于这些函数类的定义中，因为它们一般包含着检定无限多个不等式。显然，判定一个凸函数也存在同样的困难。然而，在适当的条件下，判定一个复合函数所属类别是可能的。现在较详细地研究这个问题。以下结果主要归于 Mangasarian[27]。另外，更一般的结果已由 Schaible[43,44] 得到。

定义在集合 $T\subset R^m\times R^k$ 上的实值函数 ϕ 称为在 T 上是递增-递减(增-减)的，当且仅当对任意 $(y^1, z^1)\in T$、$(y^2, z^2)\in T$ 有

$$y^2\geqslant y^1 \text{ 和 } z^2\leqslant z^1 \text{ 蕴涵 } \phi(y^2, z^2)\geqslant\phi(y^1, z^1). \tag{6.54}$$

同样，ϕ 称为在 T 上是 $\boldsymbol{y}$-递增(y-增)的，当且仅当对任意 $(y^1, z)\in T$、$(y^2, z)\in T$ 有

$$y^2\geqslant y^1 \text{ 蕴涵 } \phi(y^2, z)\geqslant\phi(y^1, z). \tag{6.55}$$

而 ϕ 称为在 T 上是 $\boldsymbol{y}$-递减(y-减)的，当且仅当对任意 $(y^1, z)\in T$、$(y^2, z)\in T$ 有

$$y^2\leqslant y^1 \text{ 蕴涵 } \phi(y^2, z)\geqslant\phi(y^1, z). \tag{6.56}$$

从这些定义即可得出下述结果。

引理 6.8

设 ϕ 是定义在开凸集 $T\subset R^m\times R^k$ 上的可微实值函数，则 ϕ 在 T 上是增-减的充分必要条件为：对每个 $(y, z)\in T$ 有

$$\nabla_y\phi(y, z)\geqslant 0,\quad \nabla_z\phi(y, z)\leqslant 0. \tag{6.57}$$

于是我们有下列定理。

定理 6.9

设 $X\subset R^n$ 是一凸集，$f(x)=(f_1(x), \cdots, f_m(x))$ 和 $g(x)=(g_1(x), \cdots, g_k(x))$ 定义在 X 上，ϕ 是在 $R^m\times R^k$ 上的实值函数。定义

$$\Phi(x)=\phi(f(x), g(x)). \tag{6.58}$$

如果下述假定中有任意一条成立：

(i) f 是凸的，g 是凹的，ϕ 是增-减的。

(ii) f 是线性的, g 是线性的.

(iii) f 是凸的, g 是线性的, ϕ 是 y-增的.

(iv) f 是凹的, g 是线性的, ϕ 是 y-减的.

那末

(a) 若 ϕ 是凸的, 则 Φ 是凸的.

(b) 若 X 是开的, f 和 g 在 X 上是可微的, ϕ 是伪凸的, 则 Φ 是伪凸的.

(c) 如果 ϕ 是拟凸的, 则 Φ 是拟凸的.

【证明】 对定理 4.12 略作修改就能证明上述部分(a). 我们现在在假定(i)下来证部分(b). 若 $x^1\in X$、$x^2\in X$, 则

$$(x^2-x^1)^T\nabla\Phi(x^1)=(x^2-x^1)^T[\nabla f(x^1)\nabla_f\phi(f(x^1),\ g(x^1))\\+\nabla g(x^1)\nabla_g\phi(f(x^1),\ g(x^1))].\qquad(6.59)$$

由(4.116)和引理 6.8, 可得

$$(x^2-x^1)^T\nabla\Phi(x^1)\leqslant[f(x^2)-f(x^1)]^T\nabla_f\phi(f(x^1),\ g(x^1))\\+[g(x^2)-g(x^1)]^T\nabla_g\phi(f(x^1),\ g(x^1)).\ (6.60)$$

因此, 如果(6.60)的左边是非负的, 则右边也是非负的; 再由 ϕ 的伪凸性可得

$$\phi(f(x^1),\ g(x^1))\leqslant\phi(f(x^2),\ g(x^2)),\qquad(6.61)$$

此即

$$\Phi(x^1)\leqslant\Phi(x^2),\qquad(6.62)$$

因而 Φ 是伪凸的. 读者能容易完成在条件(ii)、(iii)、(iv)下部分(b)的证明.

现在在假定(i)下证部分(c).

设 $x^1\in X$、$x^2\in X$, q_1、q_2 为权. 若 $\Phi(x^1)\leqslant\Phi(x^2)$, 则

$$\phi(f(x^1),\ g(x^1))\leqslant\phi(f(x^2),\ g(x^2)).\qquad(6.63)$$

因为 ϕ 是拟凸的, 有

$$\phi(q_1f(x^1)+q_2f(x^2),\ q_1g(x^1)+q_2g(x^2))\leqslant\phi(f(x^2),\ g(x^2)).\qquad(6.64)$$

再根据假定, 可得到

$$\phi(f(q_1x^1+q_2x^2),\ g(q_1x^1+q_2x^2))\leqslant\phi(f(x^2),\ g(x^2)),\quad(6.65)$$

此即

$$\Phi(q_1x^1+q_2x^2)\leqslant\Phi(x^2). \tag{6.66}$$

因而 Φ 是拟凸的，证明之余下部分是类似的. ∎

定理 6.9 能应用于一大类函数. 首先考虑**非线性分式函数**. 设 X 是 R^n 中的凸集，h_1 和 h_2 是 X 上的实值函数，又设

$$\Phi(x)=\frac{h_1(x)}{h_2(x)}. \tag{6.67}$$

设下述假定中任何一条在 X 上成立：

1. h_1 是非负且凸的；h_2 是正且凹的.
2. h_1 是非正且凹的；h_2 是负且凸的.
3. h_1 是非正且凸的；h_2 是正且凸的.
4. h_1 是非负且凹的；h_2 是负且凹的.
5. h_1 是线性的；h_2 是非零且线性的.
6. h_1 是非正且线性的；h_2 是非零且凸的.
7. h_1 是非负且线性的；h_2 是非零且凹的.
8. h_1 是凸的；h_2 是正且线性的.
9. h_1 是凹的；h_2 是负且线性的.

则 Φ 在 X 上是拟凸的. 另外，若 X 是开的，又 h_1 和 h_2 在 X 上是可微的，则 Φ 也是伪凸的.

其次，考虑**倒数函数**. 设

$$\Phi(x)=\frac{1}{h(x)}. \tag{6.68}$$

若 h 在 X 上是正且凹的，则 Φ 在 X 上是凸的. 若 h 在 X 上是负且凸的，则 Φ 在 X 上是凹的.

定理 6.9 也能应用于**双重非线性函数**. 设 $X\subset R^n$ 是开凸集，h_1 和 h_2 在 X 上是可微的，定义

$$\Phi(x)=h_1(x)h_2(x), \tag{6.69}$$

则在下述假定(1)或(2)之下，Φ 是伪凸的：

1. h_1 是非正且凸的；h_2 是正且凹的.
2. h_1 是负且凸的；h_2 是非负且凹的.

关于这些结果之扩充能在[43, 44]内找到.

R^n 上的实值二次函数之凸性，能通过检定关于它的 Hesse 矩阵的有限多个条件来验明. 然而存在仅是伪凸或拟凸而非凸的二次函数. 例如，两个变量的函数

$$g(x)=-x_1x_2 \tag{6.70}$$

对于 $x_1\geqslant 0$、$x_2\geqslant 0$ 是拟凸的，但它并不是凸的. 类似于凸的情形，为了检验二次函数的伪凸性或拟凸性，已经导出了有限检验准则，见 Cottle、Ferland[13, 14]，Ferland[16] 和 Martos[32]. 下面的讨论基于这些参考文献，结果的证明能在那里找到.

我们从几类矩阵的定义开始. 对称实矩阵 Q 和它的二次型 x^TQx 称为是**次正定的**，如果

$$x^TQx<0 \text{ 蕴涵 } Qx\geqslant 0 \text{ 或 } Qx\leqslant 0; \tag{6.71}$$

称为是**严格次正定的**，如果

$$x^TQx<0 \text{ 蕴涵 } Qx>0 \text{ 或 } Qx<0. \tag{6.72}$$

回忆一下，Q 和 x^TQx 称为是**半正定的**，如果对每个 $x\in R^n$ 有 $x^TQx\geqslant 0$，或用类似于上述两个定义的形式表示为

$$x^TQx\leqslant 0 \text{ 蕴涵 } Qx=0. \tag{6.73}$$

对于二次型 x^TQx 的凸性而言，半正定性是一个必要充分条件. 前三类矩阵定义之比较，说明半正定的二次型(平凡地)是严格次正定型，从而也是次正定的. 一个实对称矩阵，如果是(严格)次正定但并非半正定的，则称它为(严格)**仅次正定的**. 类似地，我们称函数是仅拟凸(仅伪凸)的，如果它是拟凸(伪凸)但非凸的. 于是，我们有下述基本结果.

定理 6.10

二次型 x^TQx 在非负象限上是拟凸的，当且仅当它是次正定的. 它在不包含原点的非负象限上是伪凸的，当且仅当它是严格次正定的.

这个定理的证明能在[32]中找到.

Martos 的下列结果是检定仅拟凸的二次型是否也是伪凸函数的一种简单方法.

定理 6.11

在非负象限上的仅拟凸的二次型 x^TQx 是不含原点的非负象限上的仅伪凸函数，当且仅当矩阵 Q 并不包含一行零.

现让我们转向刻划仅次正定的矩阵和二次型的特征.

定理 6.12

二次型 x^TQx 是仅次正定的充分必要条件为：

(i) 矩阵 Q 的每个元素是非正的，而且 Q 中至少存在一个负元素。

(ii) Q 的所有主子阵的行列式都是非正的。

这个定理的证明见 Cottle、Ferland[18].

最后，我们给出仅伪凸性的充分条件.

定理 6.13

设实对称矩阵 Q 的每个元素都是非正的，而且 Q 中至少存在一个负元素. 如果 Q 的所有前主子阵的行列式都是负的，则 x^TQx 是严格仅次正定的.

定理 6.12 通过主子阵的行列式提供了检验非负象限上二次型拟凸性的一种有限的检验准则. 在一定意义上，这种检验准则是关于半正定性的类似检验准则的补充. 仅次正定二次型 x^TQx 的另一有趣的性质是，Q 恰有一个负的特征值(单重)，且它对应的特征向量或者是非负的或者是非正的，但它并不是零. 我们这里简短地提一下仅拟凹性的条件.

1. 矩阵 Q 的每个元素都是非负的，且 Q 中至少存在一个正的元素.

2. 以 D_{ik} 表示 Q 的第 i 个 k 阶主子阵的行列式，则对每个 i，$(-1)^kD_{ik}$ 都是非正的.

二次函数一般是由一个二次型与一个线性函数之和组成的. 如果二次型是凸的，则上述和式也是凸的. 不幸的是伪凸或拟凸函数与线性函数之和就不一定是伪凸的或拟凸的. 然而有关二次型的以上结果已被推广到二次函数[13,14,18]，这些结果概述于下.

定理 6.14

设 f 为二次函数

$$f(x)=b^Tx+\frac{1}{2}x^TQx, \quad x\in R^n, \tag{6.74}$$

则 f 在 R^n 上是凸的充分必要条件为它是拟凸的．令

$$\bar{Q}=\begin{bmatrix} Q & b \\ b^T & 0 \end{bmatrix}, \tag{6.75}$$

则 f 在非负象限上是仅拟凸的充分必要条件为矩阵 $\bar{Q}$ 是仅次正定的．若 f 在非负象限上是仅拟凸的且 $b\neq 0$，则 f 在非负象限上也是伪凸的．

例 6.1.1[18]

考虑二次型

$$f_1(x)=-(x_1+x_2)^2, \tag{6.76}$$

写成

$$f_1(x)=x^TQx, \tag{6.77}$$

其中

$$Q=\begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix}, \tag{6.78}$$

Q 的特征值是 $\lambda^1=-2$ 和 $\lambda^2=0$，且对应 λ^1 的特征向量 $v^1=(1,1)^T$．这个二次型是非凸的，因为 Q 不是半正定的(它实际上为半负定的)．然而，根据定理 6.12，矩阵 Q 是仅次正定的．因此，根据定理 6.10，函数 f_1 在非负象限上是仅拟凸的．并且，根据定理 6.11，它还是仅伪凸的．现在考虑二次函数

$$f_2(x)=-x_1-x_2-\frac{1}{2}(x_1+x_2)^2. \tag{6.79}$$

由(6.75)，我们有

$$\bar{Q}=\begin{bmatrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & 0 \end{bmatrix}. \tag{6.80}$$

读者容易验证 $\bar{Q}$ 是仅次正定的．因此，根据定理 6.15，在非负象

限上，这个二次函数是仅拟凸和仅伪凸的． ■

显然，当凸性是充分但非必要的时候，引出诸如伪凸性或拟凸性的推广，可能获得分析和计算上的好处．如果一个数学规划中的函数是非凸的，但又“仅”属于前面讨论中曾提到过的函数类之一，这时，这些推广的研究是特别有帮助的．这样一些函数有时能用某个非线性变换把它们变成凸函数．这个问题将在下一节中研究．

在结束这一节时我们要提一下有关 Beckenbach[9]，Beckenbach、Bing[10] 的工作，在那里凸函数是用“次-Φ 函数”概念加以推广的．这些工作的主要结果也见 Berge[13]．

6.2 弧式连通集和可变换为凸的函数

我们现在以另一种观点来推广凸性．在第 4 章中，R^n 的一个子集 C 定义为凸集，如果对于 C 中任意两点，连结它们的直线段也在 C 中．这个定义能加以推广，代替直线段我们要求对于 C 中任意两点，存在一段连结它们的连续曲线全部在集合 C 内．这种观点在分析中是熟知的，能够表述如下：

集合 $S\subset R^n$ 称为是弧式连通的，如果对于每对点 $x^1\in S$、$x^2\in S$，都存在一个定义在单位区间 $[0, 1]\subset R$ 上的、取 S 中向量为值的连续函数 p_{x^1,x^2}，满足

$$p_{x^1,x^2}(0)=x^1,\quad p_{x^1,x^2}(1)=x^2, \tag{6.81}$$

就称 p_{x^1,x^2} 为连通函数或弧．注意 p_{x^1,x^2} 一般依赖于两点 x^1、x^2 及集 S．对于一个弧式连通集中的一对点 x^1 和 x^2，可以存在多于一个连通函数．对于 $x^1=x^2\in S$，我们能选取

$$p_{x^1,x^2}(\theta)=x^1,\quad 0\leqslant\theta\leqslant 1. \tag{6.82}$$

例 6.2.1

显然，凸集都是弧式连通集，而函数

$$p_{x^1,x^2}(\theta)=(1-\theta)x^1+\theta x^2,\quad 0\leqslant\theta\leqslant 1 \tag{6.83}$$

是凸集中任意两点的连通函数．现在考虑中心在原点、半径为 r 之圆的外部点的集合 $S_1\subset R^2$，即

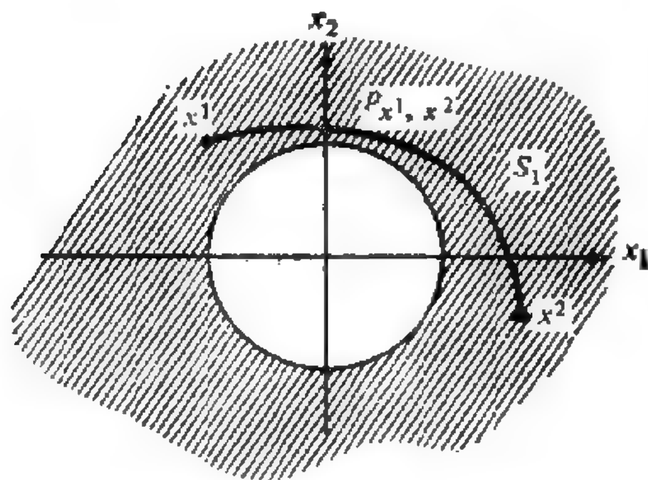


图 6.2 圆的余集是弧式连通的

$$S_1=\{(x_1, x_2):(x_1)^2+(x_2)^2\geqslant(r)^2\}. \tag{6.84}$$

于是对于任意两点 $x^1\in S_1$、$x^2\in S_1$, 用极坐标表示的下列函数是连通函数:

$$p_{x^1,x^2}(\theta)=\begin{bmatrix}[(1-\theta)r^1+\theta r^2]\cos((1-\theta)\alpha^1+\theta\alpha^2)\\ [(1-\theta)r^1+\theta r^2]\sin((1-\theta)\alpha^1+\theta\alpha^2)\end{bmatrix}, \quad 0\leqslant\theta\leqslant1, \tag{6.85}$$

其中

$$x_1^i=r^i\cos\alpha^i, \quad x_2^i=r^i\sin\alpha^i, \quad i=1, 2, \tag{6.86}$$

因而 S_1 是弧式连通集. 参看图 6.2. ▌

弧式连通集与连通集是密切相关的. 尤其是, 每个弧式连通集都是连通集, 而每个开连通集也是弧式连通集. 我们提醒读者, 集合 $D\subset R^n$ 称为是不连通的, 如果存在两个开集 A、B, 使得 $A\cap D$ 和 $B\cap D$ 是不相交的非空集合, 而且它们之和集为 D. 若 R^n 之子集不是不连通的就称为是连通的. 至于更详细的内容, 读者可参看 Apostol[1], Bartle[8], Berge[12] 或 Singer、Thorpe[45]. 图 6.3 列出了弧式连通集的一些例子.

凸函数的上图象和有效区域都是凸集, 另外, 至今讨论过的广义凸函数的定义域都是凸集. 现在我们开始研究与弧式连通集有关的凸函数的一些新的推广.

定理 4.15 表明一个函数是凸函数的充分必要条件是: 它被限

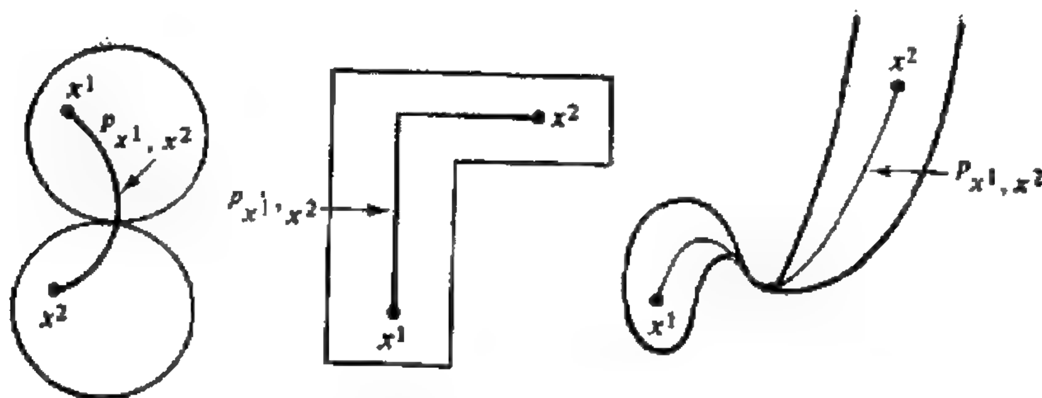


图 6.3 弧式连通集

制在定义域内每一直线段上时也是凸的. 当然，这个论断并不意味着凸函数被限制在凸集内其他弧上也是凸的. 假设我们考虑这样一类实值函数，它们定义在弧式连通集上，而且它们被限制在函数定义域内某些弧上时是凸的(若以直线段作为上述弧，普通的凸函数就成为这样一类函数). 若 S 是这样的函数的定义域(弧式连通集)，对于每对点 $x^1\in S$、$x^2\in S$，设 H_{x^1,x^2} 是具有上述性质的弧，于是我们能写

$$f(H_{x^1,x^2}(\theta))\leqslant(1-\theta)f(x^1)+\theta f(x^2),\quad 0\leqslant\theta\leqslant1. \qquad (6.87)$$

为进一步扩充这种想法，考虑这样的函数，它可以是非凸的，但能用某个严格单调的连续函数把它变换成凸函数. 事实上，对于每对实数 α^1、α^2，定义 Φ_{α^1,α^2} 为在单位区间 $[0,1]\subset R$ 上，而在 R 中取值的严格单调函数，它满足

$$\Phi_{\alpha^1,\alpha^2}(0)=\alpha^1,\quad \Phi_{\alpha^1,\alpha^2}(1)=\alpha^2. \qquad (6.88)$$

于是定义在弧式连通集 S 上的实值函数 f，称为是**弧式凸的**，如果对于每对点 $x^1\in S$、$x^2\in S$ 和 $0\leqslant\theta\leqslant1$，有

$$f(H_{x^1,x^2}(\theta))\leqslant\Phi_{f(x^1),f(x^2)}(\theta). \qquad (6.89)$$

定义实值凸函数的经典不等式(4.42)是(6.89)的特殊情形，其中，令

$$H_{x^1,x^2}(\theta)=(1-\theta)x^1+\theta x^2, \qquad (6.90)$$

$$\Phi_{f(x^1),f(x^2)}(\theta)=(1-\theta)f(x^1)+\theta f(x^2). \qquad (6.91)$$

关于弧式凸函数和有关主题的其他材料能在 Avriel、Zang[7] 和 Zang[46] 中找到.

在本节的余下部分，我们将考虑弧式凸函数的一种重要子类．如在(6.90)和(6.91)所见，在凸函数的定义中用到的 H 和 Φ 实际上分别是一对点 x^1, x^2 和它们相应的函数值 $f(x^1)$, $f(x^2)$ 的加权算术平均值．这个事实启示我们去考虑这样一类弧式凸函数，它的 H_{x^1,x^2} 和 $\Phi_{f(x^1),f(x^2)}$ 是更一般的**均值函数**[21]．我们现在更详细地考察广义凸性的这个方面．

如在凸函数情形一样，我们假定 f 是定义在 R^n 上而取值在扩充的实数集中的函数，即 $f(x)$ 或者为实数或者为 $\pm\infty$．以前所定义的概念，诸如闭、正常、有效区域等概念这里也将用到．设 h 是定义在包含 f 的有效区域在内的 R^n 的一子集上而取值在 R^n 中的连续函数，并设 h 到值域上是一对一的．同样，设 ϕ 是定义在扩充的实数集的子集上严格单调的连续函数，该定义域包含 f 的有限值域，ϕ 到它的值域上也必是一对一的．注意 h 和 ϕ 各自具有单叶的反函数 h^{-1} 和 ϕ^{-1}，满足 $h^{-1}h(x)=hh^{-1}(x)=x$ 和 $\phi^{-1}\phi(\alpha)=\phi\phi^{-1}(\alpha)=\alpha$．现假设 H_{x^1,x^2} 为 **h-均值函数**，由下式给出：

$$H_{x^1,x^2}(\theta)=h^{-1}[(1-\theta)h(x^1)+\theta h(x^2)],\quad 0\leqslant\theta\leqslant 1.\tag{6.92}$$

同样，假设 $\Phi_{f(x^1),f(x^2)}$ 为 **ϕ-均值函数**

$$\Phi_{f(x^1),f(x^2)}(\theta)=\phi^{-1}[(1-\theta)\phi f(x^1)+\theta\phi f(x^2)],\quad 0\leqslant\theta\leqslant 1.\tag{6.93}$$

从而一个在 R^n 上的正常弧式凸函数 f 称为是 **(h,ϕ)-凸的**，如果对于每对 $x^1\in R^n$, $x^2\in R^n$ 和 $0\leqslant\theta\leqslant 1$，有

$$f[h^{-1}((1-\theta)h(x^1)+\theta h(x^2))]\leqslant\phi^{-1}[(1-\theta)\phi f(x^1)+\theta\phi f(x^2)].\tag{6.94}$$

我们现在来看 (h,ϕ)-凸函数的一些例子．

例 6.2.2

(a) 首先，取 $h(x)=x$, $x\in R^n$ 和 $\phi(\alpha)=e^{r\alpha}$，其中 r 是一个非零实参数．于是代入(6.94)，得到

$$f((1-\theta)x^1+\theta x^2)\leqslant\log[(1-\theta)e^{rf(x^1)}+\theta e^{rf(x^2)}]^{\frac{1}{r}}.\tag{6.95}$$

满足(6.95)的函数也称为 **r-凸函数**，是由 Avriel[5] 和 Martos[20]

引入的.

(b) 其次,取 R^2 上的 Rosenbrock"弯谷"函数,由下式给出:

$$f(x)=100[x_2-(x_1)^2]^2+(1-x_1)^2. \tag{6.96}$$

这是一个连续可微的非凸函数,具有唯一的极小值点 $x^*=(1,1)$. 然而它并不属于前一节讨论过的任何一类广义凸函数, 如伪凸函数或拟凸函数. 事实上,如同在图 6.4 所见到的,它的水平集是非凸的"香蕉型"集合. 令

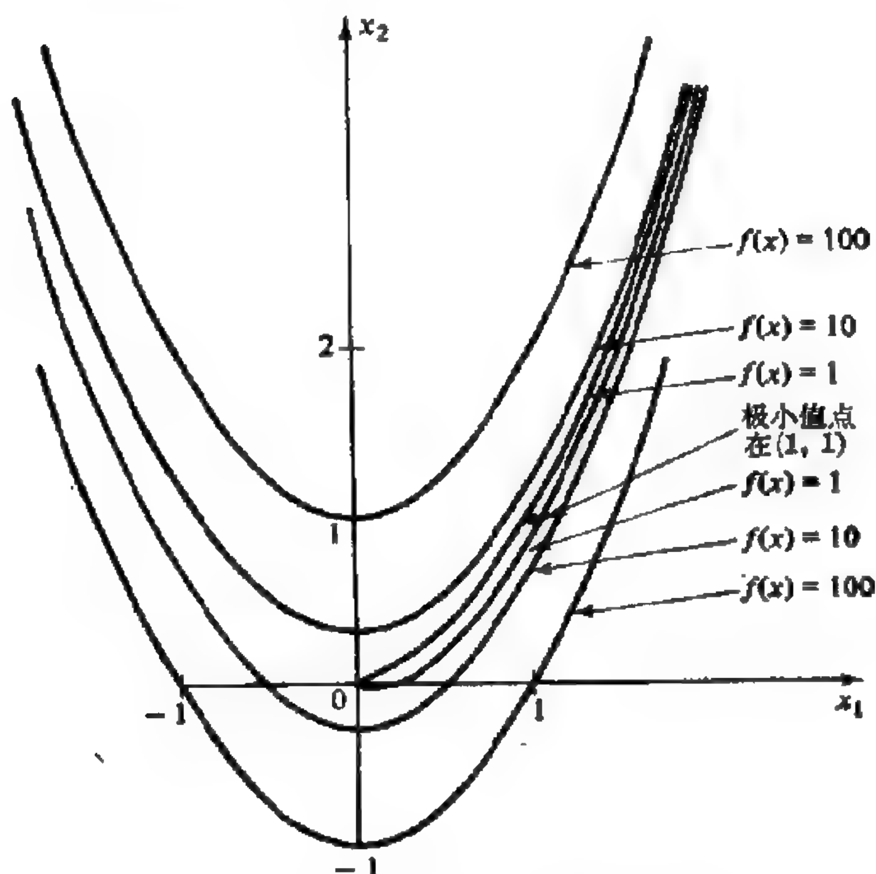


图 6.4 Rosenbrock 的"弯谷"函数

$$h(x)=\begin{pmatrix} 10[x_2-(x_1)^2] \\ 1-x_1 \end{pmatrix}, \tag{6.97}$$

$$h^{-1}(t)=\begin{pmatrix} 1-t_2 \\ \frac{1}{10}t_1+(1-t_2)^2 \end{pmatrix}, \tag{6.98}$$

又 $\phi(\alpha)=\alpha$. 对于每对 $x^1\in R^2$、$x^2\in R^2$, 我们得到

$$H_{x^1,x^2}(\theta)=\begin{pmatrix}(1-\theta)x_1^1+\theta x_1^2\\(1-\theta)[x_2^1-(x_1^1)^2]+\theta[x_2^2-(x_1^2)^2]+[(1-\theta)x_1^1+\theta x_1^2]^2\end{pmatrix}.\tag{6.99}$$

读者可以验证，对于 $0\leqslant\theta\leqslant1$，有

$$f(H_{x^1,x^2}(\theta))\leqslant(1-\theta)f(x^1)+\theta f(x^2).\tag{6.100}$$

事实上，由(6.96)和(6.98)我们得到

$$\phi f h^{-1}(t)=(t_1)^2+(t_2)^2,\tag{6.101}$$

这是一个凸的二次函数！在下面，我们将看到这个结果对每个 $(h,\ \phi)$-凸函数都成立；即它们能变换成凸函数．(a)是 h 取为恒等函数的 $(h,\ \phi)$-凸函数的例子．(b)是 ϕ 取为恒等函数的 (h,ϕ)-凸函数的例子．在下一例中，h 和 ϕ 将都不是恒等函数．

(c) 设 f 是 R^n 的正象限上的正值函数．定义 f_1 为

$$f_1(x)=\begin{cases}f(x), & x>0,\\+\infty, & \text{其他}.\end{cases}\tag{6.102}$$

对于 $x>0$，取 $h(x)=(\log x_1,\cdots,\log x_n)$．对 $\alpha>0$，取 $\phi(\alpha)=\log\alpha$．对于 f_1 的有效区域中任意两点 x^1、x^2 和 $0\leqslant\theta\leqslant1$，根据(6.94)并加以整理后，如果成立

$$f[(x_1^1)^{1-\theta}(x_1^2)^\theta,\ \cdots,\ (x_n^1)^{1-\theta}(x_n^2)^\theta]\leqslant(f(x^1))^{1-\theta}(f(x^2))^\theta,\tag{6.103}$$

那末我们定义 f_1 为(log, log)-凸函数．在非线性规划的一个称为几何规划的特殊分支中出现的正项式[16]定义为

$$g(x)=\sum_{i=1}^m c_i\prod_{j=1}^n x_j^{a_{ij}},\tag{6.104}$$

其中 $c_i>0$，$a_{ij}\in R$ 都是常量，这种函数就是(log, log)-凸函数． ∎

$(h,\ \phi)$-凸函数最重要的性质是它们可变换为凸的．换句话，正如我们将在下一个定理中所见到的，它们能被变换成凸函数．在本节余下部分，假定 ϕ 是一个严格递增函数．我们的基本结果是定理 6.15.

定理 6.15

函数 f 在 R^n 上是正常 (h, ϕ)-凸函数的充分必要条件为：由下式给出的 $\hat{f}$ 在 R^n 上是正常凸的：

$$\hat{f}(y)=\phi f h^{-1}(y). \tag{6.105}$$

【证明】 函数 f 在 R^n 上是正常 (h, ϕ)-凸的充分必要条件为：对任意 $x^1\in R^n$, $x^2\in R^n$, $0\leqslant\theta\leqslant 1$, 成立

$$f[h^{-1}((1-\theta)h(x^1)+\theta h(x^2))]\leqslant\phi^{-1}[(1-\theta)\phi f(x^1)+\theta\phi f(x^2)]. \tag{6.106}$$

因为 ϕ 是严格递增的，我们得到

$$\phi f[h^{-1}((1-\theta)h(x^1)+\theta h(x^2))]\leqslant(1-\theta)\phi f(x^1)+\theta\phi f(x^2). \tag{6.107}$$

令 $h(x)=y$, 并代入(6.107)，就得

$$\phi f h^{-1}((1-\theta)y^1+\theta y^2)\leqslant(1-\theta)\phi f h^{-1}(y^1)+\theta\phi f h^{-1}(y^2). \tag{6.108}$$

就是说 $\phi f h^{-1}$ 是正常凸的. ▌

引进广义的加法和乘法运算，可以很方便地导出 (h, ϕ)-凸函数的许多性质，这属于 Ben-Tal[11]. 设 z 为定义在 R^n 的子集 Z 上的 n 维向量值连续函数. 它具有反函数 z^{-1}. 对 $x\in Z$、$y\in Z$, 定义 **z-向量加法**为

$$x\oplus y=z^{-1}[z(x)+z(y)]. \tag{6.109}$$

对 $x\in Z$、$\lambda\in R$, 定义 **z-数乘**为

$$\lambda\odot x=z^{-1}[\lambda z(x)]. \tag{6.110}$$

类似地，设 ω 是定义在 $\Omega\subset R$ 上的连续的实值函数，它具有反函数 ω^{-1}. 则两个数 $\alpha\in\Omega$、$\beta\in\Omega$ 的 **ω-加法**定义为

$$\alpha[+]\beta=\omega^{-1}[\omega(\alpha)+\omega(\beta)]. \tag{6.111}$$

对 $\alpha\in\Omega$、$\lambda\in R$, **ω-数乘**定义为

$$\lambda[\cdot]\alpha=\omega^{-1}[\lambda\omega(\alpha)]. \tag{6.112}$$

最后，对向量 $x\in Z$、$y\in Z$, (z, ω)-内积定义为

$$(x^T y)_{z,\omega}=\omega^{-1}[(z(x))^T z(y)], \tag{6.113}$$

这里假定上式右边有意义.

从代数观点来完整地处理这些运算还需要更广泛的背景材

料．有兴趣的读者可参看[11]．

例 6.2.8

设

$$z(x)=(\log x_1,\ \cdots,\ \log x_n), \tag{6.114}$$

并设

$$Z=\{x: x\in R^n,\ x>0\}, \tag{6.115}$$

则

$$x\oplus y=[\exp(\log x_1+\log y_1),\ \cdots,\ \exp(\log x_n+\log y_n)]^T \tag{6.116}$$

$$=(x_1y_1,\ \cdots,\ x_ny_n)^T. \tag{6.117}$$

对于 $\lambda\in R$, 有

$$\lambda\odot x=[\exp(\lambda\log x_1),\ \cdots,\ \exp(\lambda\log x_n)]^T \tag{6.118}$$

$$=((x_1)^\lambda,\ \cdots,\ (x_n)^\lambda)^T. \quad \blacksquare \tag{6.119}$$

容易验证，均值函数也都可以写成这种形式，例如，

$$H_{x^1,x^2}(\theta)=[(1-\theta)\odot x^1]\oplus(\theta\odot x^2) \tag{6.120}$$

和

$$\Phi_{\alpha^1,\alpha^2}(\theta)=[(1-\theta)[\cdot]\alpha^1][+](\theta[\cdot]\alpha^2). \tag{6.121}$$

由此推知，定义正常 $(h,\ \phi)$-凸函数的不等式(6.94)可写成

$$f[((1-\theta)\odot x^1)\oplus(\theta\odot x^2)]$$
$$\leqslant((1-\theta)[\cdot]f(x^1))[+](\theta[\cdot]f(x^2)). \tag{6.122}$$

因此，$(h,\ \phi)$-凸函数在上述广义代数运算下是“凸的”．

利用这些运算，我们看一看怎样把前几章关于凸函数的结果推广到 $(h,\ \phi)$-凸函数上去．例如，定理 4.10 表明，如果 f 和 g 都是凸函数，λ 是非负数，则 $f+g$ 和 λf 也是凸函数．

相应的结果为：f 和 g 都是 $(h,\ \phi)$-凸函数，$\lambda\geqslant 0$，则 $f[+]g$ 和 $\lambda[\cdot]f$ 也是 $(h,\ \phi)$-凸的．我们已经知道，当且仅当 $-f$ 为凹时，f 是凸的．同样，当且仅当 $(-1)[\cdot]f$ 为 $(h,\ \phi)$-凹时，f 是 $(h,\ \phi)$-凸的．在第 4 章我们证明了，在凸集 $C\subset R^n$ 上的可微函数 f 是凸的充分必要条件为：对任意两点 $x^1\in C$、$x^2\in C$，有

$$f(x^2)\geqslant f(x^1)+(x^2-x^1)^T\nabla f(x^1). \tag{6.123}$$

同样，在 R^n 上的可微函数 f 是 $(h,\ \phi)$-凸的充分必要条件为

$$f(x^2) \geqslant f(x^1)[+][(x^2 \ominus x^1)^T \nabla^* f(x^1)]_{h,\phi}, \tag{6.124}$$

其中

$$x^2 \ominus x^1 = x^2 \oplus (-1) \odot x^1 = h^{-1}[h(x^2) - h(x^1)], \tag{6.125}$$

$$\nabla^* f(x^1) = h^{-1}[\nabla \phi f h^{-1}(t)|_{t=h(x^1)}]. \tag{6.126}$$

不等式(6.124)能转写成比较熟悉的通常代数运算形式，得

$$\phi f(x^2) \geqslant \phi f(x^1) + \phi'(f(x^1)) \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial f(x^1)}{\partial x_j} \frac{\partial h_j^{-1}(h(x^1))}{\partial x_i} (h_i(x^2) - h_i(x^1)). \tag{6.127}$$

当 $h(x)=x$ 时，(h, ϕ)-凸函数的这种子类的性质已由 Avriel、Zang[6] 导出，他们没有利用广义的代数运算. 那里还讨论了带有这类函数的非线性规划的某些方面.

还可以定义 (h, ϕ)-凸函数的共轭函数，并导出包含这类广义凸性的非线性规划的对偶性关系. 我们只是非常简单地提一下这个论题. 一个在 R^n 上的 (h, ϕ)-凸函数 f 的共轭函数定义为

$$f^*(\xi) = \sup_x \{(\xi^T x)_{h,\phi} [-] f(x)\} \tag{6.128}$$

或

$$f^*(\xi) = \sup_x \{\phi^{-1}[h(\xi)^T h(x) - \phi f(x)]\}, \tag{6.129}$$

这里假定上式右边有意义. 注意 (h, ϕ)-凸函数的值域可以包含 $\pm\infty$. 因而在 $R \cup \pm\infty$ 上函数的单调性必须适当地解释.

例 6.2.4

设 $x \in R$,

$$f(x) = -xe^{-x}. \tag{6.130}$$

令 $h(t)=t$, 且

$$\phi(\alpha) = \begin{cases} -\log(-\alpha), & \alpha < 0, \\ +\infty, & \alpha \geqslant 0. \end{cases} \tag{6.131}$$

可以验证 f 是 (h, ϕ)-凸的. 实际上，

$$\hat{f}(x) = \phi f h^{-1}(x) = \phi f(x) = \begin{cases} x - \log x, & x > 0, \\ +\infty, & x \leqslant 0 \end{cases} \tag{6.132}$$

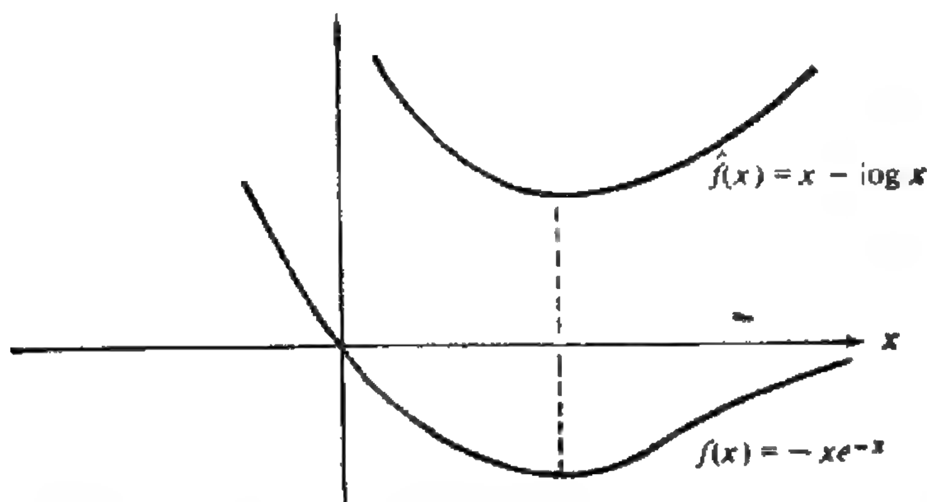


图 6.5 (h,ϕ)-凸函数 $f(x)=-xe^{-x}$ 及其变换而得的凸函数

在 R 上是凸的. 在图 6.5 中说明了这些函数. 现在, 有

$$\phi^{-1}(\alpha)=-e^{-\alpha}, \tag{6.133}$$

且

$$f^*(\xi)=\sup_x\{-\exp(-\xi x+\phi f(x))\} \tag{6.134}$$

$$=\sup_{x>0}\{-\exp(-\xi x+x-\log x)\}. \tag{6.135}$$

因此

$$f^*(\xi)=\begin{cases} e\xi-e, & \xi<1,\\ 0, & \xi\geqslant 1.\end{cases} \tag{6.136}$$

有趣的是, 注意 f^* 实际上是一个凹函数, 但它也是 (h,ϕ)-凸的, 因为

$$\hat{f}^*(\xi)=\phi f^*(\xi)=\begin{cases} -1-\log(1-\xi), & \xi<1,\\ +\infty, & \xi\geqslant 1,\end{cases} \tag{6.137}$$

而它是凸的(见图 6.6). 读者可以验证, f^* 的共轭 f^{**} 是 cl f, 恰如凸的情况一样. ∎

显然, 仅当检验这种广义凸性并不需要巨大的工作量时, 具体的 (h,ϕ)-凸函数在非线性规划的分析中才显示出有意义的应用. 因而, 问题成为如何能找到具有所需性质的函数 h 和 ϕ, 使得能够建立函数 f 的 (h,ϕ)-凸性. 至今, 对这个问题, 只有部分答案是可用的. 在某些特殊情况下, 稍作努力就能找到所需要的函数. 例

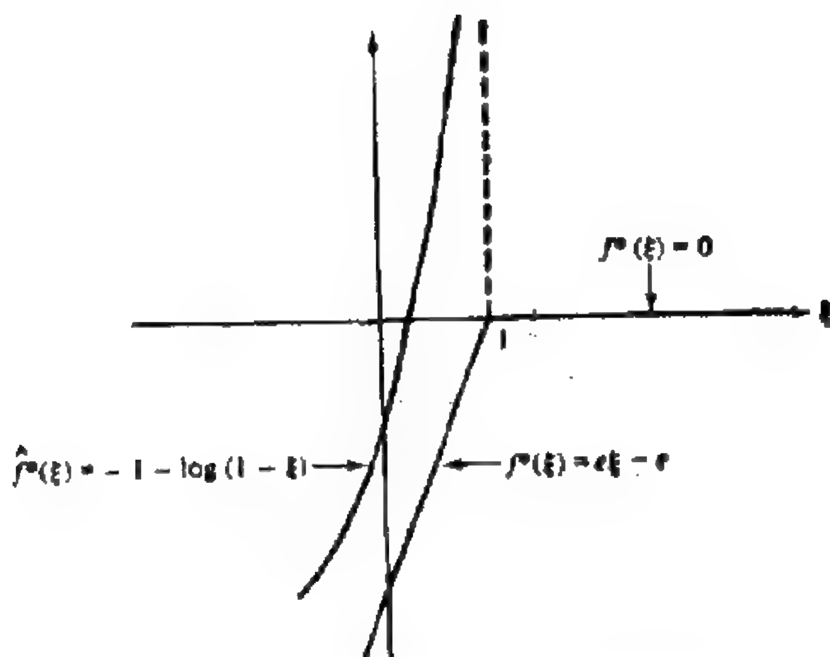


图 6.6 (h, ϕ)-凸共轭函数 f^* 及其变换而得的凸函数

如，我们已有属于 Avriel 的下述结果，它的证明能在[5]中找到。

定理 6.16

若 f 定义在开凸集 $C \subset R^n$ 上，为二次连续可微的拟凸实值函数．如果存在一实数 r^*，成立

$$r^* = \sup_{\substack{x \in C \\ \|z\|=1}} \frac{-z^T \nabla^2 f(x) z}{[z^T \nabla f(x)]^2}, \tag{6.138}$$

其中 $z^T \nabla f(x) \neq 0$，则 f 是 r^*-凸的；也就是说，当 $h(t) = t$，$\phi(\alpha) = e^{r^* \alpha}$ 时，f 为 (h, ϕ)-凸函数．

例 6.2.5

设 $f(x) = \log x$，它定义在 $C = \{x : x \in R, x > 0\}$ 上，是熟知的凹函数．我们有

$$\nabla f(x) = \frac{1}{x}, \quad \nabla^2 f(x) = -\frac{1}{(x)^2}, \tag{6.139}$$

根据(6.138)，有

$$r^* = \sup_{\substack{x > 0 \\ \|z\|=1}} \frac{[1/(x)^2](z)^2}{(1/x)^2 (z)^2} = 1. \tag{6.140}$$

因而 $\log x$ 是 1-凸的；也就是说，当 $h(t) = t$，$\phi(\alpha) = e^\alpha$ 时，f 为

(h, ϕ)-凸函数.

取 $C\subset R$ 上的函数 $f(x)=(x)^3$. 它是一个单调的递增函数, 在 $x=0$ 处有一拐点. 这里, 我们得到, 当 $x\to 0^-$ 时,

$$\sup_{\substack{x\neq 0\\ \|z\|=1}} \frac{-6x(z)^2}{9(x)^4(z)^2}=\sup_{x\neq 0}\frac{-2}{3(x)^3}=+\infty. \tag{6.141}$$

因此, 不存在实数 r^*, 使得 $(x)^3$ 是 r^*-凸函数.]

给定一个函数 f, 可以存在许多不同的 h 和 ϕ, 使得 f 是 (h, ϕ)-凸的. 我们将在练习内, 就 r-凸函数情形, 涉及这一性质. 包含本节引入的一类广义凸函数的某些非线性规划的分析, 将在下一章中讨论. 用一种不同的途径, 对于可变换为凸函数的某些类型的研究能在 Fenchel[17] 中找到.

第 4 和第 5 章介绍的有关凸函数的许多结果能推广到弧式凸函数上去. 同样, 本章第 1 节中得到的有关拟凸和其他广义凸函数的结果也能进一步推广到"弧式"-拟凸, "弧式"-伪凸和其他有关函数类上去. 这个工作留给读者作为练习.

将凸规划分析更进一步扩充到一般非线性规划上, 特别在对偶性和 Lagrange 理论领域方面, 已经取得成果. 这些结果的某些方面将在第 12 章中简短地提到. 关于详细的讨论, 读者可参看 Arrow、Gould、Howe[3], Mangasarian[36], Rockafellar[39,40,41] 和 Roode[42].

6.3 局部极小和整体极小

在这一节中, 我们介绍一种函数, 它的局部极小也是整体极小, 并讨论它的特征. 建立下列结果所用的主要工具是点到集映射的概念. 若 X 和 Y 分别为 R^p 和 R^q 中的两个集合. 如果对每一元素 $x\in X$, 指定一子集 $F(x)\subset Y$, 我们将这个对应称为 X 中点到 Y 中子集的点到集映射. 点到集映射在数学规划理论的各个分支中都要用到. 这里将不作研究. 有兴趣的读者可参看 Hogan[25] 和 Zangwill[49]. 关于点到集映射的拓扑性质, 可参看 Berge[12].

在图 6.7 中有一个点到集映射的例子，其中 $X\subset R$，而 $F(x)$ 是 R 的子集.

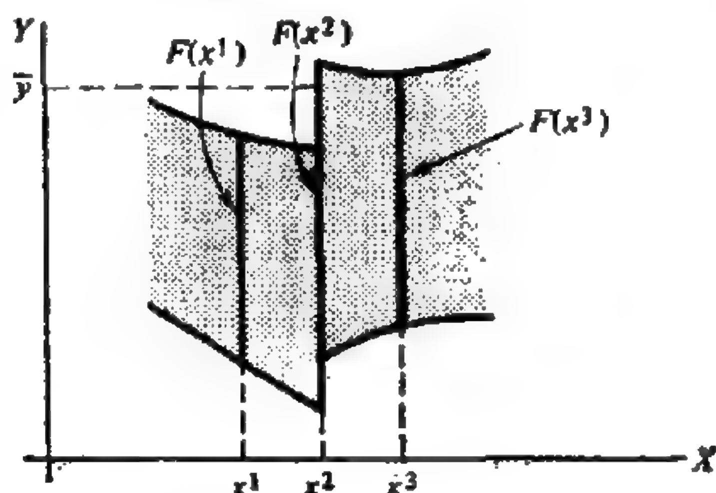


图 6.7 点到集映射

下列重要概念在后面将要用到[12]. 点到集映射 F 称为在点 $\bar{x}\in X$ 处是下半连续的 (lsc)，如果对每个 $y\in F(\bar{x})$ 和每个收敛于 $\bar{x}$ 的序列 $\{x^i\}\subset X$，存在自然数 K 和收敛于 y 的序列 $\{y^i\}$，使得

$$y^i\in F(x^i),\quad i=K,\ K+1,\ \cdots. \tag{6.142}$$

如果 F 在每一点 $x\in X$ 处都是下半连续的，则称它在 X 上是下半连续的. 图 6.7 的点到集映射 F 在 X 上不是下半连续的，因为它在 x^2 处不是下半连续的. 如果取某个 $\bar{y}\in F(x^2)$ 和一个递增的序列 $\{x^i\}$，它从 x^1 开始递增地收敛于 x^2，则不存在收敛于 $\bar{y}$ 的 $y^i\in F(x^i)$. 另一方面，易见在 x^1 或 x^3 处 F 是下半连续的.

其次考虑与实值函数相联系的一种特殊类型的点到集映射.

若 f 是在 R^n 之子集 C 上的实值函数，α 是一实数. 考察 f 的水平集

$$S(f,\ \alpha)=\{x: x\in C,\ f(x)\leqslant\alpha\}, \tag{6.143}$$

并考察集合

$$G_f=\{\alpha: \alpha\in R,\ S(f,\ \alpha)\neq\emptyset\}. \tag{6.144}$$

这样，$S(f,\ \alpha)$ 可以认为是 G_f 中的点到 R^n 的子集之点到集映射.

例 6.3.1

设 $C=R$，又

$$f(x)=(x^2). \tag{6.145}$$

于是，G_f 是 R 的非负象限.

例如，取 $\alpha=4$，于是 $\alpha\in G_f$，且

$$S(f,\ 4)=\{x: x\in R,\ |x|\leqslant 2\}. \tag{6.146}$$

下半连续性的前述定义，对于水平集映射来说，可特殊地叙述如下：

点到集映射 $S(f,\ \alpha)$ 称为在点 $\bar{\alpha}\in G_f$ 处是下半连续的，如果对每个 $x\in S(f,\ \bar{\alpha})$ 和任一收敛于 $\bar{\alpha}$ 的序列 $\{\alpha^i\}\subset G_f$，存在自然数 K 和收敛于 x 的序列$\{x^i\}$，使得

$$x^i\in S(f,\ \alpha^i),\quad i=K,\ K+1,\ \cdots. \tag{6.147}$$

如果 $S(f,\ \alpha)$ 在每一 $\alpha\in G_f$ 处是下半连续的，则称它在 G_f 上是下半连续的.

我们也可利用开邻域来定义下半连续性. 点到集映射 $S(f,\ \alpha)$ 称为在点 $\bar{\alpha}\in G_f$ 处是下半连续的，如果对于每个开集 $A\subset R^n$，只要满足

$$A\cap S(f,\ \bar{\alpha})\neq\emptyset, \tag{6.148}$$

便存在一个开邻域 $N_\delta(\bar{\alpha})$，使得对每个 $\tilde{\alpha}\in N_\delta(\bar{\alpha})\cap G_f$，都有

$$A\cap S(f,\ \tilde{\alpha})\neq\emptyset. \tag{6.149}$$

下半连续性的这两种定义在以下的定理中都要用到.

我们现在能叙述和证明关于函数的每一局部极小都是整体极小的充分必要条件，这属于 Zang、Avriel[47] 和 Zang、Choo、Avriel[48]. 先从下述充分条件开始.

定理 6.17

设 f 是在 $C\subset R^n$ 上的实值函数，$\bar{\alpha}=f(\bar{x})$，$\bar{x}\in C$. 设 $S(f,\ \alpha)$ 在 $\bar{\alpha}$ 处是下半连续的. 如果 $\bar{x}$ 是 f 的局部极小值点，那末它也必是 f 在 C 上的整体极小值点.

【证明】 如果假设条件成立，而 $\bar{x}$ 并不是 f 在 C 上的整体极小值点. 因此就存在一点 $\tilde{x}\in C$，使得

$$f(\tilde{x})<f(\bar{x}). \tag{6.150}$$

定义序列$\{\alpha^i\}$为

$$\alpha^i=\left[\frac{1}{i}f(\tilde{x})+\left(1-\frac{1}{i}\right)f(\bar{x})\right],\quad i=1,\ 2,\ \cdots. \tag{6.151}$$

显然，有

$$\lim_{i\to\infty}\{\alpha^i\}=f(\bar{x})=\bar{\alpha}, \tag{6.152}$$

且 $\bar{x}\in S(f,\ \bar{\alpha})$. 由(6.150)和(6.151)推知

$$f(\tilde{x})\leqslant\alpha^i<f(\bar{x}),\quad i=1,\ 2,\ \cdots, \tag{6.153}$$

且 $\tilde{x}\in S(f,\ \alpha^i)$, $i=1,\ 2,\ \cdots$, 因此 $\{\alpha^i\}\subset G_f$.

因为假设 $S(f,\ \alpha)$ 在 $\bar{\alpha}$ 处是下半连续的，因而存在自然数 K 和收敛于 $\bar{x}$ 的序列 $\{x^i\}$, 使得对 $i=K,\ K+1,\ \cdots$, 有 $x^i\in S(f,\ \alpha^i)$, 因此

$$f(x^i)\leqslant\alpha^i,\quad i=K,\ K+1,\ \cdots. \tag{6.154}$$

又由(6.150)得

$$f(x^i)<f(\bar{x}),\quad i=K,\ K+1,\ \cdots. \tag{6.155}$$

因为 $\{x^i\}\to\bar{x}$, 对于充分小的 $\delta>0$, 存在自然数 K_δ, 使 $x^i\in C\cap N_\delta(\bar{x})$, $i=K_\delta,\ K_\delta+1,\ \cdots$, 根据 $\bar{x}$ 为局部极小值点的假设条件，x^i 也将满足

$$f(x^i)\geqslant f(\bar{x}),\quad i=K_\delta,\ K_\delta+1,\ \cdots \tag{6.156}$$

它与(6.155)相矛盾. ▌

推论 6.18

设 f 是在 $C\subset R^n$ 上的实值函数. 如果 $S(f,\ \alpha)$ 在 G_f 上是下半连续的，则 f 的每个局部极小值点也是 f 在 C 上的整体极小值点.

现在我们证明定理 6.17 的逆定理.

定理 6.19

设 f 是在 $C\subset R^n$ 上的实值函数，$\bar{\alpha}\in G_f$, 则当下列任一假定成立时，$S(f,\ \alpha)$ 在 $\bar{\alpha}$ 处是下半连续的：

(i) 每个满足 $f(x)=\bar{\alpha}$ 的 $x\in C$, 都是 f 在 C 上的整体极小值点.

(ii) 满足 $f(x)=\bar{\alpha}$ 的任何 $x\in C$, 都不是 f 的局部极小值点.

【证明】 如果假设条件成立，而 $S(f,\ \alpha)$ 在 $\bar{\alpha}$ 处不是下半

连续的，则存在一个开集 $A\subset R^n$，满足

$$A\cap S(f,\ \bar{\alpha})\neq\emptyset, \tag{6.157}$$

且对于每个 $\delta>0$，都存在一个 $\alpha(\delta)\in N_\delta(\bar{\alpha})\cap G_f$，使得

$$A\cap S(f,\ \alpha(\delta))=\emptyset. \tag{6.158}$$

因此，我们能找到一个收敛于 $\bar{\alpha}$ 的序列 $\{\alpha^i\}\subset G_f$，使得

$$A\cap S(f,\ \alpha^i)=\emptyset,\quad i=1,\ 2,\ \cdots. \tag{6.159}$$

由此可知对所有 i，$\alpha^i<\bar{\alpha}$；否则的话，如果对某个 k 有 $\alpha^k\geqslant\bar{\alpha}$，则 $S(f,\ \bar{\alpha})\subset S(f,\ \alpha^k)$. 再由(6.157)，我们得到与(6.159)相矛盾的结果. 由此可知，对每个 $\bar{x}\in A\cap S(f,\ \bar{\alpha})$ 及每个 i，都有 $\bar{x}\notin S(f,\ \alpha^i)$；又因为 $\{\alpha^i\}\to\bar{\alpha}$，我们得到 $f(\bar{x})=\bar{\alpha}$. 此外，对每个 $x\in A$，有

$$f(x)\geqslant\bar{\alpha}=f(\bar{x}). \tag{6.160}$$

因为 A 是一个开集，由此可知，$\bar{x}$ 必定是 f 之局部极小值点，但显然 $\bar{x}$ 不是 f 在 C 上的整体极小值点. 这个结果与假设条件矛盾. ∎

推论 6.20

设 f 是 $C\subset R^n$ 上的实值函数. 如果 f 的每个局部极小值点都是 f 在 C 上的整体极小值点，那末 $S(f,\ \alpha)$ 在 G_f 上为下半连续的.

末了，作为推论 6.18 与推论 6.20 的直接结果，我们得到下列推论.

推论 6.21

设 f 是 $C\subset R^n$ 上的实值函数，f 的每个局部极小值点都是 f 在 C 上的整体极小值点的充分必要条件为：$S(f,\ \alpha)$ 在 G_f 上是下半连续的.

例 6.3.2

现在，我们来说明前面的结果. 图 6.8 中的函数在 $\hat{x}$ 处有一非整体的局部极小，而在 x^* 处有一整体极小. 又注意 $f(\hat{x})=f(\bar{x})=0$. 令 $\bar{\alpha}=0$，并取序列 $\{\alpha^i\}$，它的元素是

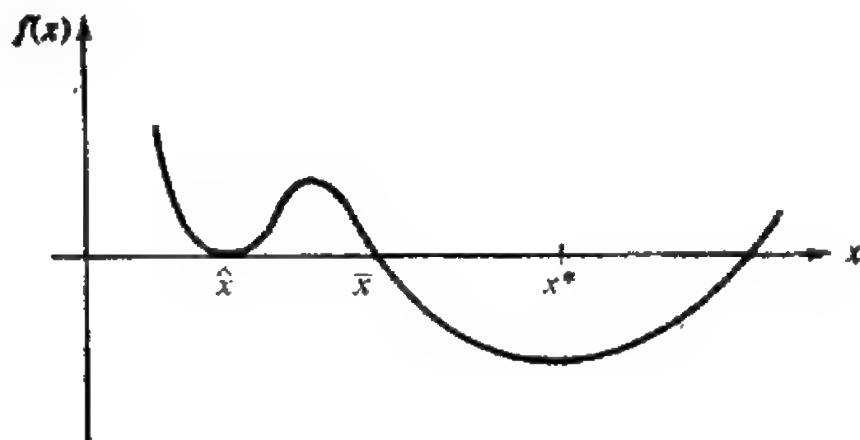


图 6.8 具有非整体的局部极小值点的函数

$$\alpha^i=\frac{1}{i}f(x^*),\quad i=1,\ 2,\ \cdots,\tag{6.161}$$

显然，$\{\alpha^i\}\subset G_f$，且$\{\alpha^i\}\to\bar{\alpha}$. 点 $\hat{x}$ 包含在 $S(f,\ 0)$ 中，但是在每个序列$\{x^i\}$中，如果对充分大的 i 有 $x^i\in S(f,\ \alpha^i)$，那末对这些 i，就必有 $x^i>\bar{x}$，于是，这样的序列均不收敛于 $\hat{x}$. 因此 $S(f,\ \alpha)$在 $\bar{\alpha}$ 处不是下半连续的. ▌

在有的情形，验证某些函数的水平集映射的下半连续性是不太困难的. 这样就可保证函数具有理想的性质：如果存在极小的话，它只有整体极小. 对于上节讨论的函数类，下面给出它的这种验证.

例 6.3.3

我们要证明，如果 f 是定义在弧式连通集 $C\subset R^n$ 上的实值$(h,\ \phi)$-凸函数，其中 C 包含由(6.92)给出的所有弧，那末 $S(f,\alpha)$ 在 G_f 上是下半连续的. 取 $\bar{\alpha}\in G_f$，并取任一收敛于 $\bar{\alpha}$ 的序列 $\{\alpha^i\}\subset G_f$. 只要考虑满足 $f(\bar{x})=\bar{\alpha}$ 的点 $\bar{x}\in S(f,\ \bar{\alpha})$就足够了，因为如果 $f(\bar{x})<\bar{\alpha}$，则对所有 i 都取 $x^i=\bar{x}$，序列$\{x^i\}$便是满足(6.147)的. 如果 $\bar{x}$ 是 f 在 C 上的一个整体极小值点，那末一定有

$$\alpha^i\geqslant\bar{\alpha},\quad i=1,\ 2,\ \cdots,\tag{6.162}$$

对每个 i 仍可选取 $x^i=\bar{x}$. 如果 $\bar{x}$ 不是 f 在 C 上的整体极小值点，那末存在一点 $\tilde{x}\in C$ 使得

$$f(\tilde{x})<f(\bar{x})=\bar{\alpha},\tag{6.163}$$

定义序列$\{\hat{x}^k\}$如下：

$$\hat{x}^k=h^{-1}\left[\frac{1}{k}h(\tilde{x})+\left(1-\frac{1}{k}\right)h(\bar{x})\right],\quad k=1,2,\cdots. \tag{6.164}$$

显然，$\{\hat{x}^k\}$收敛于$\bar{x}$，且$\{\hat{x}^k\}\subset C$。我们不妨假定，对每个k有$\|\hat{x}^k-\bar{x}\|<1/k$。

由f的(h,ϕ)-凸性推知，对$k=1,2,\cdots$有

$$\begin{aligned}f(\hat{x}^k)&=f\left[\left(\frac{1}{k}\odot\tilde{x}\right)\oplus\left(\left(1-\frac{1}{k}\right)\odot\bar{x}\right)\right]\\&\leqslant\left[\frac{1}{k}[\cdot]f(\tilde{x})\right][+]\left[\left(1-\frac{1}{k}\right)[\cdot]f(\bar{x})\right].\end{aligned} \tag{6.165}$$

由(6.163)和(6.165)，我们得到

$$f(\hat{x}^k)\leqslant f(\bar{x})=\bar{\alpha}. \tag{6.166}$$

要完成这个证明，我们还需构造一个序列$\{x^i\}$，它应具有$S(f,\alpha)$的下半连续性所要求之性质。由(6.166)，并由$\{\alpha^i\}$收敛于$\bar{\alpha}$，推知存在K_1使得

$$f(\hat{x}^1)\leqslant\alpha^i,\quad i=K_1,K_1+1,\cdots. \tag{6.167}$$

对于序列$\{\alpha^i\}$中每个α^i，$i=K_1+1,\cdots$，令x^i是用下列方法定义的。选择$\bar{k}(i)$，使得

$$\bar{k}(i)=\sup\{k:f(\hat{x}^k)\leqslant\alpha^i\}, \tag{6.168}$$

这里假定上述有限上确界存在；否则令$\bar{k}(K_1)=1$，且

$$\bar{k}(i)=\bar{k}(i-1)+1. \tag{6.169}$$

此外，令

$$x^i=\begin{cases}\hat{x}^1, & i=1,\cdots,K_1,\\ \hat{x}^{\bar{k}(i)}, & i=K_1+1,\cdots.\end{cases} \tag{6.170}$$

我们现在证明$\{x^i\}$收敛于$\bar{x}$。即对任何正数λ，存在自然数$K(\lambda)$，使得

$$x^i\in N_\lambda(\bar{x}),\quad i=K(\lambda),K(\lambda)+1,\cdots. \tag{6.171}$$

定义

$$\delta(k)=\frac{1}{k},\quad k=1,2,\cdots, \tag{6.172}$$

和

$$\hat{k}(\lambda)=\begin{cases}\max\left\{k:\delta(k-1)=\dfrac{1}{k-1}>\lambda\right\}, & \lambda<1,\\ 1, & \lambda\geqslant 1,\end{cases} \tag{6.173}$$

则 $N_{\delta(\hat{k}(\lambda))}(\bar{x})\subset N_\lambda(\bar{x})$，又

$$f(\hat{x}^{\hat{k}(\lambda)})<\bar{\alpha}, \tag{6.174}$$

于是，我们能找到 $\overline{K}(\lambda)$，满足 $\overline{K}(\lambda)\geqslant K_1$ 和 $\overline{K}(\lambda)\geqslant\hat{k}(\lambda)$，且使得

$$f(\hat{x}^{\hat{k}(\lambda)})\leqslant\alpha^i, \quad i=\overline{K}(\lambda),\ \overline{K}(\lambda)+1,\ \cdots. \tag{6.175}$$

从而，可能出现下列情况之一：

1．存在一个 $\bar{i}\geqslant\overline{K}(\lambda)$，对于它，$\bar{k}(\bar{i})$ 由(6.168)得到．这时对于 $i\geqslant\bar{i}$，由(6.168)、(6.169)和(6.175)，得到 $\bar{k}(i)\geqslant\hat{k}(\lambda)$ 和

$$\hat{x}^{\bar{k}(i)}\in N_{\delta(k(i))}(\bar{x})\subset N_{\delta(\hat{k}(i))}(\bar{x})\subset N_\lambda(\bar{x}). \tag{6.176}$$

因此

$$x^i\in N_\lambda(\bar{x}), \quad i=\bar{i},\ \bar{i}+1,\ \cdots. \tag{6.177}$$

2．对所有 $i\geqslant\overline{K}(\lambda)$，序列 $\{x^i\}$ 的元素是根据 (6.169) 和 (6.170) 选取的．显然，$i\geqslant\overline{K}(\lambda)$，$x^i$ 是从收敛于 $\bar{x}$ 的序列$\{\hat{x}^k\}$中依次取得的．

因此在这两种情况下，$\{x^i\}$都收敛于 $\bar{x}$，而且

$$x^i\in S(f,\ \alpha^i), \quad i=K_1,\ K_1+1,\ \cdots. \tag{6.178}$$

练　　习

6．A． 证明拟凸函数可有另一种定义：对任何 α，集合

$$S'(f,\ \alpha)=\{x:x\in X\subset R^n,\ f(x)<\alpha\} \tag{6.179}$$

都是凸的．〔提示：应用定理 4.2．〕把它与定理 6.1 相比较．

6．B． 定义在 $X\subset R^n$ 上的实值函数 f 称为是拟单调的[81]，如果 f 既是拟凸又是拟凹的．证明，若 f 是拟单调的，则对每个 $\alpha\in R$，

$$\bar{S}(f,\ \alpha)=\{x:x\in X,\ f(x)=\alpha\} \tag{6.180}$$

都是凸集．反之，证明：如果对每个 $\alpha\in R$，$\bar{S}(f,\ \alpha)$都是凸的，而且 f 是连续的，则 f 是拟单调的．以线性分式函数 g 说明这些结果，g 定义在凸集 $C\subset R^n$ 上，由下式给出：

$$g(x)=\frac{a^Tx+b}{c^Tx+d}, \tag{6.181}$$

其中分母永不为零.

6.C. 众所周知，线性函数在一凸子集 $X\subset R^n$ 上的极值必在 X 的顶点处达到(如果极值存在). 推广这个结果到伪线性函数，即既是伪凸又是伪凹的函数. 上一练习中所定义的线性分式函数，是否是伪线性的?

6.D. 证明，如果二次函数

$$f(x)=b^Tx+\frac{1}{2}x^TQx,\quad x\in R^n, \tag{6.182}$$

是仅拟凸的，则 $b\leqslant 0$.

6.E. (a) 确定下列矩阵的正定性、次正定性或负定性等:

$$(\mathrm{i})\begin{bmatrix}1&1&2\\1&1&2\\2&2&0\end{bmatrix},\quad(\mathrm{ii})\begin{bmatrix}0&-1&-1\\-1&0&-1\\-1&-1&0\end{bmatrix},\quad(\mathrm{iii})\begin{bmatrix}1&1&1\\1&0&0\\1&0&0\end{bmatrix}. \tag{6.183}$$

(b) 设

$$Q=\begin{bmatrix}-1&-1\\-1&0\end{bmatrix},\quad b=(0,-2)^T, \tag{6.184}$$

$f(x)=b^Tx+\frac{1}{2}x^TQx$, f 在非负象限中是否为拟凸的?

6.F. (a) 设 $X=\{x:x\in R,\ 0\leqslant x\leqslant 1\}$, $f(x)=-(x)^2-x$ 定义在 X 上. 证明 f 在 X 上是严格伪凸的. 它是不是强拟凸的?

(b) 现在，设 $f(x)=-(x)^2$. 证明 f 在同一 X 上不是伪凸的. 它是不是拟凸的?

6.G. (a) Ortega、Rheinboldt[86] 定义伪凸函数如下: 在凸集 $X\subset R^n$ 上的实值函数 f 称为伪凸的，如果对任意的 $x^1\in X$、$x^2\in X$，满足

$$f(x^1)>f(x^2), \tag{6.185}$$

都存在正数 α 和 τ, $\tau\leqslant 1$(一般 α 和 τ 都依赖于 x^1 和 x^2)，使得对每个 $0\leqslant\theta\leqslant\tau$ 满足

$$f((1-\theta)x^1+\theta x^2)\leqslant f(x^1)-\theta\alpha. \tag{6.186}$$

证明，对于可微函数，上述定义等价于本章所给出的定义.

(b) 如果 $x^1\neq x^2$ 且 $f(x^1)\geqslant f(x^2)$ 时 (6.186) 都成立，则 Ortega、Rheinboldt 称 f 为严格伪凸的. 另一方面如果

$$x^1\neq x^2 \text{ 和 } f(x^1)\geqslant f(x^2) \text{ 蕴涵 } (x^2-x^1)^T\nabla f(x^1)<0, \tag{6.187}$$

Ponstein[87] 定义此可微函数 f 是严格伪凸的. 证明这两个定义对可微函数是等价的. 证明函数

$$f(x)=\begin{cases}x, & x<0,\\ (x)^2+1, & x\geqslant 0,\end{cases} \tag{6.188}$$

是在 Ortega-Rheinboldt 意义下严格伪凸的.

6.H. 证明连续的实值函数在 R^n 上是严格拟凸的充分必要条件为：它的所有的水平集都是凸的，且关于使得 $f(x)=\alpha$ 的点 $x\in R^n$ 的集合 $\bar{S}(f,\alpha)$，它或者被包含在 $S(f,\alpha)$ 的边界点集合中，或者 $\bar{S}(f,\alpha)=S(f,\alpha)$。

6.I. 证明：如果 f 是 R^n 上连续的实值拟凸函数，而且它的局部极小都是整体极小，则 f 是严格拟凸的．在证明中可以应用前面练习的结果．

6.J. 通过推广(4.119)和(4.144)，导出关于可微 (h,ϕ)-凸函数的另外的定义．

6.K. 若 r 和 s 是实数，$s>r$。又若 $\xi_1,\cdots,\xi_m$ 都是正数．众所周知[21]，成立*)

$$\left\{\sum_{i=1}^{m} q_i(\xi_i)^r\right\}^{1/r} \leqslant \left\{\sum_{i=1}^{m} q_i(\xi_i)^s\right\}^{1/s}. \tag{6.189}$$

证明：如果 f 是 r-凸函数，则对任意 $s>r$，f 也是 s-凸函数．利用关系式

$$\lim_{r\to+\infty}\left\{\sum_{i=1}^{m} q_i(\xi_i)^r\right\}^{1/r}=\max(\xi_1,\cdots,\xi_m), \tag{6.190}$$

证明每个 r-凸函数也是拟凸的．

6.L. 证明：在 R^n 上两次连续可微的实值函数 f 是 r-凸的充分必要条件为：矩阵 Q 即

$$Q(x)=\nabla^2 f(x)+r\nabla f(x)[\nabla f(x)]^T, \tag{6.191}$$

对每个 $x\in R^n$ 都是半正定的．

6.M. 给定在 R^2 上的函数

$$f(x)=[x_2-(x_1)^3+(4-x_1)^2]^{1/2}, \tag{6.192}$$

寻找 h 和 ϕ，使得 f 是 (h,ϕ)-凸的．

6.N. 证明：如果 f 是在 R^n 的凸子集上的严格拟凸函数，则 $S(f,\alpha)$ 是 G_f 上的下半连续的点到集映射．

参 考 文 献

1. Apostol, T. M., *Mathematical Analysis*, 2nd ed., Addison-Wesley Publishing Co., Reading, Mass., 1974.

2. Arrow, K. J., and A. C. Enthoven, "Quasi-Concave Programming," *Econometrica*, 29, 779–800 (1961).

3. Arrow, K. J., F. J. Gould, and S. M. Howe, "A General Saddle Point Result for Constrained Optimization," *Math. Prog.*, 5, 225–234 (1973).

*) 译注：q_i 为权，即满足 $q_i\geqslant 0$，$\sum q_i=1$.

4. Arrow, K. J., L. Hurwicz, and H. Uzawa, "Constraint Qualifications in Maximization Problems," *Naval Res. Log. Quart.*, **8**, 175–191 (1961).

5. Avriel, M., "r-Convex Functions," *Math. Prog.*, **2**, 309–323 (1972).

6. Avriel, M., and I. Zang, "Generalized Convex Functions with Applications to Nonlinear Programming," Chapter 2 in *Mathematical Programs for Activity Analysis*, P. Van Moeseke (Ed.), North-Holland Publishing Co., Amsterdam, 1974.

7. Avriel, M., and I. Zang, "Generalized Convex Sets and Functions in Nonlinear Programming," in preparation.

8. Bartle, R. G., *The Elements of Real Analysis*, 2nd ed. John Wiley & Sons, New York, 1976.

9. Beckenbach, E. F., "Generalized Convex Functions," *Bull. Am. Math. Soc.*, **43**, 363–371 (1937).

10. Beckenbach, E. F., and R. H. Bing, "On Generalized Convex Functions," *Bull. Am. Math. Soc.*, **51**, 220–230 (1945).

11. Ben-Tal, A., "On Generalized Means and Generalized Convexity," *J. Optimization Theory & Appl.*, to appear.

12. Berge, C., *Topological Spaces*, Oliver & Boyd Ltd., Edinburgh, 1963.

13. Cottle, R. W., and J. A. Ferland, "Matrix-Theoretic Criteria for Quasi-Convexity and Pseudo-Convexity of Quadratic Functions," *Linear Alg. Appl.*, **5**, 123–136 (1972).

14. Cottle, R. W., and J. A. Ferland, "On Pseudo-Convex Functions of Nonnegative Variables," *Math. Prog.*, **1**, 95–101 (1971).

15. de Finetti, B., "Sulle Stratificazioni Convesse," *Ann. Math. Pura Appl.*, **30**, 173–183 (1949).

16. Duffin, R. J., E. L. Peterson, and C. Zener, *Geometric Programming—Theory and Applications*, John Wiley & Sons, New York, 1967.

17. Fenchel, W., *Convex Cones, Sets and Functions*, Mimeographed lecture notes, Princeton University, Princeton, N.J., 1951.

18. Ferland, J. A., "Quasi-Convex and Pseudo-Convex Functions on Solid Convex Sets," Ph.D. dissertation, Stanford University, Stanford Calif., 1971.

19. Greenberg, H. J., and W. P. Pierskalla, "A Review of Quasi-Convex Functions," *Operations Research*, **19**, 1553–1570 (1971).

20. Hanson, M. A., "Bounds for Functionally Convex Optimal Control Problems," *J. Math. Anal. & Appl.*, **8**, 84–89 (1964).

21. Hardy, G. H., J. E. Littlewood, and G. Pólya, *Inequalities*, 2nd ed., Cambridge University Press, Cambridge, England, 1952.

22. Hogan, W. W., "Point-to-Set Maps in Mathematical Programming," *SIAM Review*, **15**, 591–603 (1973).

23. Karamardian, S., "Duality in Mathematical Programming," *J. Math. Anal. & Appl.*, **20**, 344–358 (1967).

24. LUENBERGER, D. G., "Quasi-Convex Programming," *SIAM J. Appl. Math.*, **16**, 1090–1095 (1968).

25. MANGASARIAN, O. L., "Pseudo-Convex Functions," *J. SIAM Control, Ser. A*, **3**, 281–290 (1965).

26. MANGASARIAN, O. L., *Nonlinear Programming*, McGraw-Hill Book Co., New York, 1969.

27. MANGASARIAN, O. L., "Convexity, Pseudo-Convexity and Quasi-Convexity of Composite Functions," *Cahiers du Centre d'Etude de Rech. Oper.*, **12**, 114–122 (1970)

28. MANGASARIAN, O. L., "Unconstrained Lagrangians in Nonlinear Programming," *SIAM J. Control*, **13**, 772–791 (1975).

29. MARTOS, B., "Hyperbolic Programming," *Naval Res. Log. Quart.*, **11**, 135–155 (1964).

30. MARTOS, B., "Nem-Lineáris Programmozási Módszerek Hatóköre (The Power of Nonlinear Programming Methods)," MTA Közgazdaságtudományi Intézetének Közleményei No. 20, Budapest, 1966 (in Hungarian).

31. MARTOS, B., "Quasi-Convexity and Quasi-Monotonicity in Nonlinear Programming," *Studia Sci. Math. Hung.*, **2**, 265–273 (1967).

32. MARTOS, B., "Subdefinite Matrices and Quadratic Forms," *SIAM J. Appl. Math.*, **17**, 1215–1223 (1969).

33. MEYER, R., "The Validity of a Family of Optimization Methods," *SIAM J. Control*, **8**, 41–54 (1970).

34. MINCH, R. A., "Applications of Symmetric Derivatives in Mathematical Programming," *Math. Prog.*, **1**, 307–320 (1971).

35. NIKAIDÔ, H., "On von Neumann's Minimax Theorem," *Pacific J. Math.*, **4**, 65–72 (1954).

36. ORTEGA, J. M., and W. C. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

37. PONSTEIN, J., "Seven Kinds of Convexity," *SIAM Review*, **9**, 115–119 (1967).

38. ROCKAFELLAR, R. T., *Convex Analysis*, Princeton University Press, Princeton, N.J., 1970.

39. ROCKAFELLAR, R. T., "Penalty Methods and Augmented Lagrangians in Nonlinear Programming," in *Proceedings of the 5th IFIP Conference on Optimization Techniques*, R. Conti and A. Ruberti (Eds.), Springer-Verlag, Berlin, 1973.

40. ROCKAFELLAR, R. T., "Augmented Lagrange Multiplier Functions and Duality in Nonconvex Programming," *SIAM J. Control*, **12**, 268–285 (1974).

41. ROCKAFELLAR, R. T., "A Dual Approach to Solving Nonlinear Programming Problems by Unconstrained Optimization," *Math. Prog.*, **5**, 354–373 (1973).

42. ROODE, J. D., "Generalized Lagrangian Functions in Mathematical Programming," Doctoral dissertation, University of Leiden, The Netherlands, 1968.

43. SCHAIBLE, S., "Beiträge zur quasikonvexen Programmierung," Doctoral dissertation, University of Cologne, Germany, 1971.

44. SCHAIBLE, S., "Quasiconvex Optimization in General Real Linear Spaces," *Zeitschrift für Operations Research*, **16**, 205–213 (1972).

45. SINGER, I. M., and J. A. THORPE, *Lecture Notes on Elementary Topology and Geometry*, Scott, Foresman and Co., Glenview, Ill., 1967.

46. ZANG, I., "Generalized Convex Programming," D.Sc. dissertation, Technion, Israel Institute of Technology, Haifa, 1974.

47. ZANG, I., and M. AVRIEL, "On Functions whose Local Minima are Global," *J. Optimization Theory & Appl.*, **16**, 183–190 (1975).

48. ZANG, I., E. U. CHOO, and M. AVRIEL, "A Note on Functions whose Local Minima are Global," *J. Optimization Theory & Appl.*, **18**, 555–559 (1976).

49. ZANGWILL, W. I., *Nonlinear Programming: A Unified Approach*, Prentice-Hall, Englewood Cliffs, N.J., 1969.

第 7 章

几种非线性规划问题的分析

在结束第 I 部分的这一章中，我们选择了三种具有代表性的非线性规划，用以说明前几章的方法和结果。这些规划已经广泛地由许多学者用不同的方法进行研究和分析。我们在分析时，一般将遵循那些已发表在有关著作中的结果，至少在用语上是如此。但是，在多数情况下，读者能够根据前几章的内容，无困难地证明这里所介绍的那些结果。

在第 1 节中我们论述二次规划，从分析和计算的观点来看，它也许是最接近线性规划的一类非线性规划。在第 2 节中，提出了某种类型的随机规划问题，将它列入的理由之一，是提醒读者注意存在这种规划的表述形式，这点对三类规划问题均相同。随机规划代表着一类重要的非线性最优化问题，许多专著和教科书中都已将有关内容写入。在最后一节中我们叙述几何规划的某些背景，特别是导出了它的对偶规划。由于几何规划及其推广在实际应用方面有着明显的巨大潜力，近年来它已成为许多著作的主题。

遗憾的是，因本书的篇幅限制，不能更广泛地讨论这三种类型问题和某些应予列入的其他问题，我们只能建议有兴趣的读者进一步参看更专门的数学规划文献。

7.1 二次规划

考虑非线性规划

(PQP) $$\min f(x)=a+c^Tx+\frac{1}{2}x^TQx \tag{7.1}$$

受限制于

$$A^Tx\geqslant b, \tag{7.2}$$

这里 $a\in R$、$c\in R^n$、$b\in R^m$ 是给定的向量，Q 是 $n\times n$ 对称实矩阵，A 是 $n\times m$ 实矩阵. 规划(PQP)称为二次规划. 为简单起见，假定 $m<n$，A 的秩为 m. 显然，目标函数和约束函数都是二次连续可微的. 而且，约束函数

$$g_i(x)=(a^i)^T x-b_i,\quad i=1,\ \cdots,\ m \tag{7.3}$$

的梯度是线性无关的，这里 a^i 是 A 的第 i 列. 因此，关于最优点必要条件的一阶和二阶约束品性在每一能行点均成立. 与(PQP)相联系的 Lagrange 式为

$$L(x,\ \lambda)=a+c^T x+\frac{1}{2}x^T Qx-\sum_{i=1}^{m}\lambda_i[(a^i)^T x-b_i]. \tag{7.4}$$

现在叙述由定理 3.8 和定理 3.10 导出的最优点必要条件.

定理 7.1

设 x^* 是问题(PQP)的一个解，则存在一个向量 $\lambda^*\in R^m$，使得

$$c+Qx^*-A\lambda^*=0, \tag{7.5}$$

$$\lambda_i^*[(a^i)^T x^*-b_i]=0,\quad i=1,\ \cdots,\ m, \tag{7.6}$$

$$\lambda^*\geqslant 0. \tag{7.7}$$

同时，设 $z\neq 0$ 满足

$$z^T a^i=0,\quad i\in I(x^*), \tag{7.8}$$

其中

$$I(x^*)=\{i:(a^i)^T x^*=b_i,\ i=1,\ \cdots,\ m\}, \tag{7.9}$$

则必有

$$z^T Qz\geqslant 0. \tag{7.10}$$

转到充分条件，我们首先定义

$$\hat{I}(x^*)=\{i:i\in I(x^*),\ \lambda_i^*>0\}. \tag{7.11}$$

于是由定理 3.11，我们有下列定理.

定理 7.2

设 x^* 对问题(PQP)是能行的. 如果存在向量 λ^* 满足(7.5)至(7.7)，且对满足下列各式的 $z\neq 0$：

$$z^T a^i=0,\quad i\in \hat{I}(x^*), \tag{7.12}$$

$$z^T a^i\geqslant 0,\quad i\in I(x^*),\ i\notin \hat{I}(x^*), \tag{7.13}$$

必有

$$z^TQz>0, \tag{7.14}$$

那么 x^* 是(PQP)的一个严格局部极小值点.

因为目标函数是二次可微的,而且约束是线性的.规划(PQP)的凸的性质完全能由矩阵 Q 刻划.

在第4章和第6章,我们叙述过使 f 为凸、伪凸和拟凸的关于 Q 的条件,我们只考虑最简单的情况.在这一节的其余部分,假定 Q 是正定的.由此可知,f 是严格凸的,并且,出现在前面两个定理中关于 z^TQz 的所有二阶条件必定满足.这样,一阶条件(7.5)至(7.7)是必要且充分的.

我们还能叙述一些对偶性结果.首先,我们引出对应于(PQP)的对偶规划.仿照第5章的推导,我们把(PQP)变换为具有扰动的"无约束"规划:

$$\min \phi(x, w)=\begin{cases} f(x), & A^Tx-b\geqslant w, \\ +\infty, & \text{其他}. \end{cases} \tag{7.15}$$

这里 $x\in R^n$、$w\in R^m$. ϕ 的共轭函数由下式给出:

$$\phi^*(\xi, \lambda)=\sup_{x, w}\{\xi^Tx+\lambda^Tw-\phi(x, w)\} \tag{7.16}$$

$$=\sup_{A^Tx-b\geqslant w}\left\{\xi^Tx+\lambda^Tw-a-c^Tx-\frac{1}{2}x^TQx\right\}, \tag{7.17}$$

且可得

$$\phi^*(\xi, \lambda)=\begin{cases} \sup\limits_{x}\left\{(\xi+A\lambda-c)^Tx-\frac{1}{2}x^TQx-b^T\lambda-a\right\}, & \lambda\geqslant 0, \\ +\infty, & \text{其他}. \end{cases} \tag{7.18}$$

(7.18)中严格凹函数的上确界能通过微分来求,我们得到

$$\phi^*(\xi, \lambda)=\begin{cases} \frac{1}{2}(\xi+A\lambda-c)^TQ^{-1}(\xi+A\lambda-c)-b^T\lambda-a, & \lambda\geqslant 0, \\ +\infty, & \text{其他}. \end{cases} \tag{7.19}$$

(PQP)的对偶规划成为

(DQP) $$\max v(\lambda, \eta)=b^T\lambda-\frac{1}{2}\eta^TQ^{-1}\eta+a \tag{7.20}$$

受限制于

$$A\lambda-\eta=c, \tag{7.21}$$

$$\lambda\geqslant 0. \tag{7.22}$$

这样，(DQP)也是一个二次规划. 注意，引入向量变数 η 只是为了记法上的方便.

让我们考察这对对偶二次规划的某些性质. 原有的扰动函数由下式给出:

$$\Phi(w)=\inf_x\{f(x):A^Tx-b\geqslant w\}, \tag{7.23}$$

这里将空集的下确界理解为 $+\infty$ 以及

$$\Phi(0)=\inf_x\{f(x):A^Tx-b\geqslant 0\}. \tag{7.24}$$

因此，若(PQP)存在一个能行点 x, 则 $\Phi(0)<+\infty$, 而且(PQP)通常是稳定的. 对 (DQP) 也能作类似的分析. (DQP) 的能行集决不是空集, 因为对任何 $\lambda\geqslant 0$, 我们总能找到满足(7.21)的 η. 然而，(DQP)的最优值为无界的情形则可能发生. 为了看到这一点，取一个 $n<m$ 的 $n\times m$ 矩阵 A, 比如 $A=[-1,\ 1]$, 向量 $b=(1,1)^T$, 且 $Q=1, c=a=0$, 则由对偶目标函数求极大:

$$\max -\phi^*(0, \lambda)=\max_{\lambda\geqslant 0}\left\{b^T\lambda-\frac{1}{2}(A\lambda-c)^TQ^{-1}(A\lambda-c)+a\right\}, \tag{7.25}$$

选择 $\lambda_1=\lambda_2$ 二者同时趋于 $+\infty$, 我们得到 $-\phi^*(0, \lambda)\to+\infty$. 显然，(PQP)没有能行解且 $\Phi(0)\to+\infty$; 这就是说(PQP)是不稳定的. 从 5.2 节结果我们得到下列定理.

定理 7.8

对(PQP)的任何能行解 x 和(DQP)的任何能行解 (λ, η), 我们有

$$f(x)\geqslant v(\lambda, \eta). \tag{7.26}$$

规划 (PQP) 是稳定的且有一个最优解 x^* 的充要条件为: (DQP) 是稳定的且有一个最优解 (λ^*, η^*). 而且，若两个规划都

有最优解，则

$$f(x^*)=v(\lambda^*, \eta^*). \tag{7.27}$$

应该提及，Cottle[7]，Dennis[12] 和 Dorn[13,14,15] 等首先建立了二次规划的对偶性定理，他们没有使用这里所用的共轭函数理论．这些早期结果的重要性，主要在于它们对于对偶性理论中更新、更一般的结果的发展有着促进作用．

如同我们在 5.3 节所见，最优性条件紧密地联系着 Lagrange 式 对于这里的规划，我们有下列定理．

定理 7.4

如果 L 是由(7.4)给出的 Lagrange 式，(x^*, λ^*) 是 L 关于一切 x 和 $\lambda\geqslant 0$ 的一个鞍点，那末 x^* 是 (PQP) 的一个最优解，λ^* 是 (DQP)的一个最优解．并且还成立

$$f(x^*)=L(x^*, \lambda^*)=v(\lambda^*, \eta^*), \tag{7.28}$$

这里 η^* 满足(7.21)．反之，如果(PQP)是稳定的且有最优解 x^*，那末存在一个 λ^*，使(x^*, λ^*)是 L 的一个鞍点．对偶地，若(DQP)是稳定的且 λ^* 是一最优解，则存在一个 x^* 使(x^*, λ^*)是 L 的一个鞍点．

二次规划问题在实践中常常发生，它或者代表某些经济模型，或者作为求解更一般非线性规划中的一个子问题．关于它们的解，已经提出了许多数值解法，其中之一在第 13 章讨论．有兴趣的读者也可以查阅 Boot[6]，Dantzig[11] 的书或 Beale[5]，Cottle[8]，Cottle、Mylander[10]，Keller[21] 的文章和那里引用的参考文献．

7.2 带有可分离偿付函数的随机线性规划

下面将习惯形式的线性规划作为一个活动分析模型来考虑．我们涉及的商品以 1, …, m 为标号，活动以 1, …, n 为标号．假设 a_{ij} 表示在活动 j 中用去($a_{ij}>0$ 时)或生产($a_{ij}<0$ 时)商品 i 的数量，并设 p_j 是活动 j 的单位收益．那末，当每个活动掌握在水平 x_j 时，数量

$$y_i=\sum_{j=1}^{n} a_{ij}x_j \tag{7.29}$$

将表示用去的 $(y_i>0)$ 或生产的 $(y_i<0)$ 商品 i 的全部数量．供应量或需求量用 $b=(b_1, \cdots, b_m)^T$ 表示，规定了 b 之后，就得到一个活动分析模型的普通线性规划，可写为：

$$\max f(x)=p^Tx \tag{7.30}$$

受限制于

$$Ax=b, \tag{7.31}$$

$$x\geqslant 0. \tag{7.32}$$

在随机线性规划中 b 被认为是一个随机变量，因此，提法及其分析等新问题就随之产生．这里我们将论述偿付型的表述形式，它们的详细内容能够在 Walkup、Wets[27]，Wets[28] 以及 Williams[29] 中找到．我们的分析基本上按照 Avriel、Williams[3,4] 的处理，以共轭函数理论为基础．

我们假定，必须作出一种决策，也就是说，首先选择 x．这一步引出了某个 y 和收益 p^Tx．下一步，b 的实际数值显现了，要采取某个校正措施去消除 y 和显现值 b 之间的差异．假设我们已给出一个依赖于向量 $b-y$ 的可分离的偿付函数 g，它使得

$$g(b-y)=\sum_{i=1}^{m} g_i(b_i-y_i). \tag{7.33}$$

这个偿付函数当 $g>0$ 时指明了附加支出，当 $g<0$ 时指明了收入．于是，**具有可分离偿付的随机线性规划**在于选择第一步的 x，使整体期望收益达到最大．形式地，我们得到

$$\text{(PRP)} \qquad \max_{x\geqslant 0} p^Tx-\sum_{i=1}^{m} E[g_i(t_i-y_i)], \tag{7.34}$$

其中 E 是随机变量 t_i 的数学期望运算，y_i 由(7.29)给出．对每个 y_i，假定 $E[g_i(t_i-y_i)]$ 是有限的．我们对问题(PRP)的分析将包括它的存在性、特性和对偶性定理．假设 g_i 是 y_i 的正常凸函数，我们就要在非负象限上求一个正常凹函数的极大．因为规划(PRP)通过明显的符号改变可变为一个凸规划，所以能把第 4 章和第 5 章的结果运用于我们的分析．然而，在此之前，我们先推导

几个更一般的结果.

考虑凹非线性规划

$$\text{(CP)} \qquad \max \psi(x) = u(x) - z(Ax), \tag{7.35}$$

其中 $x \in R^n$, u 是 R^n 上的闭的正常凹函数, A 是 $m \times n$ 实矩阵, z 是 R^m 上的闭的正常凸函数. 为了计算(7.35)的对偶规划, 我们先引入一个扰动向量 $w \in R^m$, 从而得到

$$\psi(x, w) = u(x) - z(Ax - w). \tag{7.36}$$

下一步计算凹共轭函数 ψ^*:

$$\psi^*(\xi, \lambda) = \inf_{x, w} \{\xi^T x + \lambda^T w - \psi(x, w)\} \tag{7.37}$$

$$= \inf_{x, w} \{\xi^T x + \lambda^T w - u(x) + z(Ax - w)\}. \tag{7.38}$$

作代换

$$w = Ax - y, \tag{7.39}$$

我们得到

$$\psi^*(\xi, \lambda) = \inf_{x, y} \{\xi^T x - \lambda^T y + \lambda^T Ax - u(x) + z(y)\} \tag{7.40}$$

$$= \inf_{x} \{(\xi + A^T\lambda)^T x - u(x)\} - \sup_{y} \{\lambda^T y - z(y)\} \tag{7.41}$$

和

$$\psi^*(\xi, \lambda) = u^*(\xi + A^T\lambda) - z^*(\lambda). \tag{7.42}$$

这样, (7.35)的对偶规划由下式给出:

$$\text{(CD)} \qquad \min\{-\psi^*(0, \lambda)\} = z^*(\lambda) - u^*(A^T\lambda). \tag{7.43}$$

与这些规划相联系, 可以定义另一个问题(它不是一个最优化问题):

(CC) 找一个 $x \in R^n$ 和 $\lambda \in R^m$ 使

$$A^T\lambda \in \partial u(x), \tag{7.44}$$

$$Ax \in \partial z^*(\lambda), \tag{7.45}$$

其中 $\partial u(x)$ 和 $\partial z^*(\lambda)$ 分别是 u 和 z^* 在 x 处和 λ 处的次微分.

这三个问题用下面一些定理彼此联系起来[23].

定理 7.5

为使向量 x^* 和 λ^* 满足

$$u(x^*)-z(Ax^*)=\max\{u(x)-z(Ax)\} \tag{7.46}$$
$$=\min\{z^*(\lambda)-u^*(A^T\lambda)\}$$
$$=z^*(\lambda^*)-u^*(A^T\lambda^*), \tag{7.47}$$

其充分必要条件是 x^* 和 λ^* 满足

$$A^T\lambda^*\in\partial u(x^*),\quad Ax^*\in\partial z^*(\lambda^*). \tag{7.48}$$

这个定理的证明由与第 5 章类似的论证导出.

对于前面三个问题，我们有一个存在性定理.

定理 7.6

假设在 ψ 和 ψ^* 的有效区域内分别存在向量 x 和 λ，若(CP)和(CD)二者都是稳定的，则(CP)、(CD)和(CC)必定有解.

下面，给出一个刻划特性的定理.

定理 7.7

假设(CP)是正规的. 一对向量(x^*, λ^*)是(CC)的解的充要条件为: x^* 在(CP)中为最优以及 λ^* 在(CD)中为最优.

最后，我们叙述强对偶性定理.

定理 7.8

规划(CP)是稳定的且有一个最优解的充要条件为：规划(CD)是稳定的且有一个最优解. 在这个情况下，ψ 在(CP)的极大值等于 $-\psi^*$ 在(CD)的极小值.

在特殊情况 $A=I$ (恒等矩阵) 下的对偶性定理由 Fenchel[17] 首先研究，以后由 Rockafellar[23] 推广和加强.

对于可分离的函数 u 和 z

$$u(x)=\sum_{j=1}^{n}u_j(x_j), \tag{7.49}$$

$$z(y)=\sum_{i=1}^{m}z_i(y_i), \tag{7.50}$$

其共轭函数也是可分离的(见练习 5.E). 因此我们有一对可分离的对偶规划

(PSP) $$\max\sum_{j=1}^{n}u_j(x_j)-\sum_{i=1}^{m}z_i((Ax)_i) \tag{7.51}$$

和

(DSP) $$\min \sum_{i=1}^{m} z_i^*(\lambda_i) - \sum_{j=1}^{n} u_j^*((A^T\lambda)_j). \tag{7.52}$$

现在转到带可分离偿付的随机规划，注意到它是如同(7.51)给出的一个可分离规划．令

$$u_j(x_j) = \begin{cases} p_j x_j, & x_j \geqslant 0, \\ -\infty, & x_j < 0, \end{cases} \quad j=1, \cdots, n, \tag{7.53}$$

$$z_i(y_i) = E[g_i(t_i - y_i)], \quad i=1, \cdots, m, \tag{7.54}$$

我们得到

$$u_j^*((A^T\lambda)_j) = \begin{cases} 0, & (A^T\lambda)_j \geqslant p_j, \\ -\infty, & (A^T\lambda)_j < p_j, \end{cases} \tag{7.55}$$

$$z_i^*(\lambda_i) = \sup_{y_i} \{\lambda_i y_i - E[g_i(t_i - y_i)]\}. \tag{7.56}$$

于是，对偶偿付规划成为

(DRP) $$\min_{\lambda} \sum_{i=1}^{m} \sup_{y_i} \{\lambda_i y_i - E[g_i(t_i - y_i)]\} \tag{7.57}$$

受限制于

$$A^T\lambda \geqslant p. \tag{7.58}$$

注意，由假定知 $\mathrm{ED}(E[g_i]) = R$，即整个实轴．

让我们稍为细致地分析一下对偶目标函数(7.57)．其中关于 y_i 的上确界能确切地计算的充要条件为

$$\lambda_i \in \partial E[g_i(t_i - y_i)]. \tag{7.59}$$

因为其中的函数的定义域是 R，这个条件等价于条件

$$D^+ E[g_i(t_i - y_i; 1)] \geqslant \lambda_i \geqslant D^- E[g_i(t_i - y_i; 1)]. \tag{7.60}$$

可以证明

$$D^+ E[g_i(t_i - y_i; 1)] = -E[D^- g_i(t_i - y_i; 1)], \tag{7.61}$$

$$D^- E[g_i(t_i - y_i; 1)] = -E[D^+ g_i(t_i - y_i; 1)]. \tag{7.62}$$

因此，(7.60)能写为

$$-E[D^- g_i(t_i - y_i; 1)] \geqslant \lambda_i \geqslant -E[D^+ g_i(t_i - y_i; 1)]. \tag{7.63}$$

由于这些 g_i 是正常凸函数，$-g_i$ 是正常凹函数，因此，$-D^+ g_i(t_i - y_i; 1)$ 和 $-D^- g_i(t_i - y_i; 1)$ 关于 $t_i - y_i$ 是非增的，也就是关于 y_i 非减．当 $y_i \to +\infty$（或 $-\infty$）时，左右导数的极限是相等的，并且可能

取无限值. 假设

$$\lim_{y_i\to+\infty}\{-D^+g_i(t_i-y_i;1)\}=\lim_{y_i\to+\infty}\{-D^-g_i(t_i-y_i;1)\}=\gamma_i,\quad i=1,\cdots,m,\qquad(7.64)$$

$$\lim_{y_i\to-\infty}\{-D^+g_i(t_i-y_i;1)\}=\lim_{y_i\to-\infty}\{-D^-g_i(t_i-y_i;1)\}=\delta_i,\quad i=1,\cdots,m.\qquad(7.65)$$

因此对任何 y_i 和 t_i, 有

$$\gamma_i\geqslant-D^-g_i(t_i-y_i;1)\geqslant-D^+g_i(t_i-y_i;1)\geqslant\delta_i,\quad i=1,\ \cdots,\ m,\qquad(7.66)$$

且对任何 y_i, 成立

$$\gamma_i\geqslant-E[D^-g_i(t_i-y_i;1)]\geqslant-E[D^+g_i(t_i-y_i;1)]\geqslant\delta_i,\ i=1,\cdots,m.\qquad(7.67)$$

于是,(7.63)中的界限蕴涵着

$$\gamma_i\geqslant\lambda_i\geqslant\delta_i,\qquad(7.68)$$

并对 $i=1,\ \cdots,\ m$ 有

$$\lambda_i=\begin{cases}\delta_i, & \text{仅当对某个 } y_i \text{ 有 } \delta_i=-E[D^+g_i(t_i-y_i;1)],\\ \gamma_i, & \text{仅当对某个 } y_i \text{ 有 } \gamma_i=-E[D^-g_i(t_i-y_i;1)].\end{cases}\qquad(7.69)$$

我们现在能够叙述带可分离偿付的随机线性规划的基本的存在性定理.

定理 7.9

规划(PRP)恒有能行解. 若存在一个 λ 满足 (7.58)、(7.68) 和(7.69), 且(DRP)是稳定的, 则(PRP)有一个最优解.

我们还有一个刻划(PRP)的特性的定理.

定理 7.10

假设(PRP)是稳定的. 非负向量 x^* 是(PRP)的最优解的充要条件为: 存在某个 λ^* 满足

$$A^T\lambda^*\geqslant P,\qquad(7.70)$$

$$\gamma\geqslant\lambda^*\geqslant\delta\qquad(7.71)$$

和

$$\lambda_i^* = \begin{cases} \delta_i, \text{ 仅当对某个 } y_i \text{ 有 } \delta_i = -E[D^+ g_i(t_i - y_i; 1)], \\ \gamma_i, \text{ 仅当对某个 } y_i \text{ 有 } \gamma_i = -E[D^- g_i(t_i - y_i; 1)], \end{cases} \tag{7.72}$$

还有

$$-E[D^- g_i(t_i - (Ax^*)_i; 1)] \geqslant \lambda_i^* \geqslant -E[D^+ g_i(t_i - (Ax^*)_i; 1)], \quad i = 1, \cdots, m, \tag{7.73}$$

而且

$$x_j^* > 0 \tag{7.74}$$

仅仅出现在下式成立的时候:

$$\sum_{i=1}^m a_{ij} \lambda_i^* = p_j. \tag{7.75}$$

要求读者证明(7.70)至(7.75)是问题(CC)的条件(对带有偿付的随机规划来表达的),问题(CC)在(7.44)和(7.45)中给出.

最后,我们叙述强对偶性定理.

定理 7.11

规划(PRP)有一最优解 x^* 的充要条件为:规划(DRP)有一个最优解 λ^*,且在这个情形下,

$$p^T x^* - \sum_{i=1}^m E[g_i(t_i - (Ax^*)_i)] = \sum_{i=1}^m \sup_{y_i} \{\lambda_i^* y_i - E[g_i(t_i - y_i)]\}. \tag{7.76}$$

我们来更精确地规定偿付函数 g_i. 对这些偿付函数,可以有许多不同的假设. 但我们将只提最简单的偿付形式,即**线性偿付**的情况,这时,对 $i = 1, \cdots, m$,

$$-g_i(b_i - y_i) = \begin{cases} \gamma_i (b_i - y_i), & y_i \geqslant b_i, \\ \delta_i (b_i - y_i), & y_i \leqslant b_i, \end{cases} \tag{7.77}$$

这里,向量 $\gamma = (\gamma_1, \cdots, \gamma_m)^T$、$\delta = (\delta_1, \cdots, \delta_m)^T$ 以及 $\gamma_i \geqslant \delta_i$ 是给定的. Williams[29] 证明了,带有线性偿付的随机线性规划由下式给出:

$$(\text{PLRP}) \max_{x \geqslant 0} p^T x - \delta^T Ax - \sum_{i=1}^m \left\{ (\gamma_i - \delta_i) \left[\int_{-\infty}^{(Ax)_i} F_i(t) dt \right] - \delta_i E(b_i) \right\}, \tag{7.78}$$

其中,F_i 是概率

$$F(t)=\text{prob}\{b\leqslant t\} \tag{7.79}$$

的边缘分布.

要求读者导出对应于(PLRP)的对偶规划为

$$\text{(DLRP)}\quad \min\sum_{i=1}^{m}\left\{(\lambda_i-\delta_i)\hat{y}_i-(\gamma_i-\delta_i)\left[\int_{-\infty}^{\hat{y}_i}F_i(t)dt\right]+\delta_i E(b_i)\right\} \tag{7.80}$$

受限制于

$$\gamma\geqslant\lambda\geqslant\delta, \tag{7.81}$$

$$A^T\lambda\geqslant p, \tag{7.82}$$

这里 $\hat{y}_i$ 满足(F_i^+ 和 F_i^- 的含义见下面)

$$[\gamma_i F_i^+(\hat{y}_i)+\delta_i(1-F_i^+(\hat{y}_i))]\geqslant\lambda_i\geqslant[\gamma_i F_i^-(\hat{y}_i)+\delta(1-F_i^-(\hat{y}_i))]. \tag{7.83}$$

由于 F_i 可以是不连续的,对于给定的一个数 τ, 方程

$$F_i(y_i)=\tau \tag{7.84}$$

即使 $0<\tau<1$ 也可能无解. 根据这个理由,我们引入记号

$$F^+(\hat{y})=\inf_{y\to\hat{y}^+}F(y), \tag{7.85}$$

$$F^-(\hat{y})=\sup_{y\to\hat{y}^-}F(y), \tag{7.86}$$

则

$$F_i^+(\hat{y}_i)\geqslant\tau\geqslant F_i^-(\hat{y}_i) \tag{7.87}$$

对于 $0\leqslant\tau\leqslant1$ 总是有解 $\hat{y}_i$.

带有可分离偿付的随机规划的对偶性定理在这些问题的数值解法中是有用的, 因为它们对近似的线性规划的最优解提供了某种界限,并指出近似解与真实最优值的偏差. 至于详节,读者可参考 Williams[30], 那里讨论了线性偿付的情况. Ziemba[63] 以不同的处理提出了这种问题的解法. 对随机规划的一般领域有兴趣的读者, 可以在 Sengupta[25], Vajda[26] 的书中找到说明材料, 也可以在文献里的许多评论中找到这种材料.

7.3 几何规划

在非线性规划中,从分析和计算的观点出发,最广泛地被研究

的一个分支是几何规划和它的推广. 对这类非线性规划的巨大兴趣有两个方面: 一方面, 为非线性规划导得的几乎每一个分析结果和计算方法都可以用几何规划问题很好地加以阐明; 另一方面, 来自不同学科的实际最优化问题都能表述成几何规划或它的推广. 在这里我们介绍几何规划的某些分析情况, 它们说明了前几章中得到的理论结果. 对这类非线性规划有兴趣的读者, 可以在文献中找到涉及几何规划的大量论文. 作为这个论题的引导, 推荐 Duffin、Peterson、Zener 的书[16] 以及 Zener 的书[82]. 至于更广泛的论述, 参看 Avriel、Rijckaert、Wilde[1], 那里还介绍了较近的结果.

为了研究几何规划, 我们从定义**正项式** g 着手, 它是如下有限项之和组成的实函数:

$$g(x)=\sum_i c_i\prod_{j=1}^n (x_j)^{a_{ij}}, \tag{7.88}$$

这里 $c_i>0$ 和 $a_{ij}\in R$ 是给定的常数, 且 $x>0$. 一般地说, 正项式既非凸的, 也非凹的, 然而我们即将看到, 它们是(log, log)-凸函数.

设* $h(t)=\log t$, $h^{-1}(y)=e^y$ 和 $\phi(t)=\log t$, 我们得到

$$\hat{g}(y)=\phi g h^{-1}(y)=\log\sum_i c_i \exp\sum_{j=1}^n a_{ij}y_j. \tag{7.89}$$

现在由定理 6.15, 当且仅当 $\phi g h^{-1}$ 是凸函数时, 实函数 g 为(h,ϕ)-凸函数. 我们将利用下述著名的 Hölder 不等式[19,20] 来证明(7.89)中函数的凸性: 设 $\alpha_i>0$, $\beta_i>0$ 是正数, 且设 $p>0$、$r>0$ 使得 $(p)^{-1}+(r)^{-1}=1$, 则

$$\sum_i \alpha_i\beta_i\leqslant\left(\sum_i(\alpha_i)^p\right)^{1/p}\left(\sum_i(\beta_i)^r\right)^{1/r}. \tag{7.90}$$

引理 7.12

设 g 是定义在 R^n 的正象限上的一个正项式, 则由(7.89)给出的 $\hat{g}(y)$ 是 R^n 上的凸函数.

【证明】 设 $y^1\in R^n$、$y^2\in R^n$ 是任意两个点, 并且令

---

* 译注: 应为 $h(t)=(\log t_1,\cdots,\log t_n)$, $h^{-1}(y)=(e^{y_1},\cdots,e^{y_n})$.

$$\alpha_i=\left\{c_i\exp\sum_{j=1}^{n}a_{ij}y_j^1\right\}^{q_1},\quad \beta_i=\left\{c_i\exp\sum_{j=1}^{n}a_{ij}y_j^2\right\}^{q_2},\tag{7.91}$$

$$(p)^{-1}=q_1,\qquad (r)^{-1}=q_2.\tag{7.92}$$

代入(7.90)，得到

$$\sum_i\left\{c_i\exp\sum_{j=1}^{n}a_{ij}y_j^1\right\}^{q_1}\left\{c_i\exp\sum_{j=1}^{n}a_{ij}y_j^2\right\}^{q_2}$$
$$\leqslant\left(\sum_i\left\{c_i\exp\sum_{j=1}^{n}a_{ij}y_j^1\right\}\right)^{q_1}\left(\sum_i\left\{c_i\exp\sum_{j=1}^{n}a_{ij}y_j^2\right\}\right)^{q_2}.\tag{7.93}$$

两边取对数便导出

$$\log\sum_i c_i\exp\sum_{j=1}^{n}a_{ij}(q_1y_j^1+q_2y_j^2)$$
$$\leqslant q_1\log\sum_i c_i\exp\sum_{j=1}^{n}a_{ij}y_j^1+q_2\log\sum_i c_i\exp\sum_{j=1}^{n}a_{ij}y_j^2.\tag{7.94}$$

从而，ϑ 是凸函数. ▌

现在我们可以定义几何规划. 下列形式的非线性规划称为几何规划:

(PGP) $$\min g_0(x)\tag{7.95}$$

受限制于

$$g_k(x)\leqslant 1,\quad k=1,\ \cdots,\ m,\tag{7.96}$$

$$x>0,\tag{7.97}$$

这里 $g_0,\ g_1,\ \cdots,\ g_m$ 是正项式. 于是，可设

$$g_k(x)=\sum_{i\in I_k}c_i\prod_{j=1}^{n}(x_j)^{a_{ij}},\tag{7.98}$$

其中

$$I_k=\{r_k,\ \cdots,\ s_k\},\quad k=0,\ 1,\ \cdots,\ m,\tag{7.99}$$

$$r_0=1,\tag{7.100}$$

$$r_k=s_{k-1}+1,\qquad k=1,\ \cdots,\ m,\tag{7.101}$$

$$s_m=M.\tag{7.102}$$

因此(PGP)由 n 个正变数的一个目标函数和 m 个正项式约束组成. 第 k 个正项式是 s_k-r_k+1 个称为单项式的项之和. 在这规划中单项式的总数是 M.

因为 x 限定为正，我们能作变量代换

$$x_j = e^{y_j}, \quad j=1, \cdots, n, \tag{7.103}$$

并取 $g_0, g_1, \cdots, g_m$ 的对数(不改变极值点位置的单调变换)，得到一个凸规划，它服从第 4 章和第 5 章[19] 提出的分析. 在分析几何规划时，纵然这种途径为大多数作者所使用，但我们这里采取直接的(和等价的)途径，把(PGP)看作(log, log)-凸规划，从而在用于对偶性时可以阐明第 6 章的某些结果.

在介绍我们的结果之前，必须略述凸对偶性扩展到(h, ϕ)-凸规划的相应推广.

需回忆的重要事实是，利用前面一章所定义的推广的代数运算和某些容易的修正，所有凸对偶性的结果也能应用于更一般的情形.

一个标准的(h, ϕ)-凸规划给出为[31]

(SHP)
$$\min_x f(x) \tag{7.104}$$

受限制于

$$g_i(x) \geqslant 0_\phi, \quad i=1, \cdots, m, \tag{7.105}$$

其中 f 和 $g_1, \cdots, g_m$ 分别是 R^n 上的正常 (h, ϕ)-凸函数和正常 (h, ϕ) 凹函数，且

$$0_\phi = \phi^{-1}(0). \tag{7.106}$$

R^n 上的一个(h, ϕ)-凸函数称为是正常的，如果 $\mathrm{ED}(f) \neq \emptyset$ 且对每个 $x \in R^n$ 有 $f(x) > -\infty_\phi$，其中

$$\mathrm{ED}(f) = \{x : x \in R^n, f(x) < +\infty_\phi\}, \tag{7.107}$$

$$+\infty_\phi = \lim_{y \to +\infty} \phi^{-1}(y), \quad -\infty_\phi = \lim_{y \to -\infty} \phi^{-1}(y). \tag{7.108}$$

在(7.108)中的极限也可以是无限. 例如，若 $\phi(t) = \log t$，则

$$+\infty_\phi = \lim_{y \to +\infty} e^y = +\infty, \ -\infty_\phi = \lim_{y \to -\infty} e^y = 0. \tag{7.109}$$

现在引入(SHP)中向量 $w = (w_1, \cdots, w_l)^T$ 形式的扰动，使 $f(x, w)$ 和 $g(x, w)$ 分别是$((h, v), \phi)$-凸函数和$((h, v), \phi)$-凹函数. 也就是说，$v_1(w_1), \cdots, v_l(w_l)$ 是均值函数，其逆函数为 $v_1^{-1}(u_1), \cdots, v_l^{-1}(u_l)$，它们分别使 $\phi f(h^{-1}(y), v^{-1}(u))$ 和

$\phi g_i(h^{-1}(y), v^{-1}(w))$ 是通常的凸函数和凹函数. 类似于 (5.36), 我们定义

$$\theta(x, w)=\begin{cases} f(x, w), & (x, w)\in \mathrm{ED}(f),\ g_i(x, w)\geqslant 0_\phi, \\ & i=1, \cdots, m, \\ +\infty_\phi & \text{其他}. \end{cases} \tag{7.110}$$

于是可知, 通过适当选择 w 参与有关函数的方式, (SHP) 能等价于 $\theta(x, 0_v)$ 关于 x 求极小.

我们前已看到, 共轭函数是在凸情况下导出对偶规划的主要工具之一. 共轭函数的概念在第 6 章中对 (h, ϕ)-凸函数作了进一步的推广. 回想起当 f 是 (h, ϕ)-凸函数时, 在 (6.128) 中给出的定义是

$$f^*(\xi)=\sup_x\{(\xi^T x)_{h,\phi}[-]f(x)\}. \tag{7.111}$$

容易证明, f^* 也是 (h, ϕ)-凸函数, 且 $f^{**}=\mathrm{cl}\, f$, 此处闭包运算涉及闭的上图象. 我们能对一个给定的 f, 定义无限的一族共轭函数. 设一个广义的内积为

$$(x^T y)_{h^1, h^2, \phi}=\phi^{-1}[(h^1(x))^T h^2(y)], \tag{7.112}$$

这里 h^1 和 h^2 不必是相同的均值函数. 若 $h^1=h^2=h$, 我们得到

$$(x^T y)_{h, h, \phi}=(x^T y)_{h,\phi}, \tag{7.113}$$

这里右边曾在 (6.113) 中定义. 推广 (6.128) 和 (6.129), 能把一个 (h, ϕ)-凸函数 f 的共轭函数写为

$$f^*(\xi)=\sup_x\{(\xi^T x)_{\bar h, h,\phi}[-]f(x)\} \tag{7.114}$$

或

$$f^*(\xi)=\sup_x\{\phi^{-1}[\bar h(\xi)^T h(x)-\phi f(x)]\}, \tag{7.115}$$

这里再次指出, h 和 $\bar h$ 不必是相同的均值函数.

还可以证明, 由 (7.114) 定义的 f^* 是 $(\bar h, \phi)$-凸函数, 并且

$$f^{**}(x)=\sup_\xi\{(x^T\xi)_{h,\bar h,\phi}[-]f^*(\xi)\}=\mathrm{cl}\, f. \tag{7.116}$$

于是, 对应于 (SHP) 的一对原有-对偶规划是

$$(\mathrm{P}_\theta) \qquad \min_x \theta(x,\ 0_v) \tag{7.117}$$

和

$$(\mathrm{D}_{\theta^*}) \qquad \max_\lambda [-]\theta^*(0_{\bar{\xi}},\ \lambda) = \eta(\lambda), \tag{7.118}$$

这里 θ^* 是 θ 的共轭函数，由下式定义：

$$\theta^*(\xi,\ \lambda) = \sup_{x,w}\{((\xi^T x)_{\bar{h},h,\phi}[+](\lambda^T w)_{\bar{v},v,\phi})[-]\theta(x, w)\}, \tag{7.119}$$

并且，

$$[-]\theta^* = \phi^{-1}[-\phi(\theta^*)]. \tag{7.120}$$

注意(7.118)中函数 η 是 λ 的$(\bar{v},\ \phi)$-凹函数.

我们先用例题说明一个较为简单的情况，然后回到几何规划，那里要利用上述一般的一对对偶规划.

例 7.3.1

考虑问题

$$\min f(x) = (x)^{1/3} \tag{7.121}$$

受限制于

$$g(x) = (x)^{1/6} \geqslant 1, \tag{7.122}$$

这里 $x \in R$. 注意，f 实际上是凹函数，能行集是$\{x: x \in R,\ x \geqslant 1\}$. 最优解显然是 $x^* = 1$ 且 $f(x^*) = 1$. 然而，容易证明，对所有 $t \in R$，令 $h(t) = t$ 和 $\phi(t) = t^3 - 1$，就得到一个标准的$(h,\ \phi)$-凸规划：

$$\min (x)^{1/3} \tag{7.123}$$

受限制于

$$(x)^{1/6} \geqslant 0_\phi. \tag{7.124}$$

注意 $\phi^{-1}(y) = (y+1)^{1/3}$ 和 $0_\phi = 1$，$+\infty_\phi = +\infty$. 在 (7.122) 中引入扰动变量 w 及 $v(t) = t$，我们得到

$$\theta(x,\ w) = \begin{cases} (x)^{1/3}, & (x^{1/3} - w)^{1/3} \geqslant 0_\phi, \\ +\infty_\phi, & \text{其他}. \end{cases} \tag{7.125}$$

下面，我们取 $\bar{h}(t) = h(t) = t$ 来计算 θ^*.

$$\theta^*(\xi, \lambda) = \sup_{x, w}\{(((\xi x)_{\lambda,\phi}[+](\lambda w)_{v,\phi})[-]\theta(x, w)\} \quad (7.126)$$

$$= \sup_{x, w}\{\phi^{-1}[\xi x + \lambda w - \phi\theta(x, w)]\} \quad (7.127)$$

$$= \sup_{((x)^{1/3}-w)^{1/3}\geq 0_\phi}\{\xi x + \lambda w - x + 2\}^{1/3}. \quad (7.128)$$

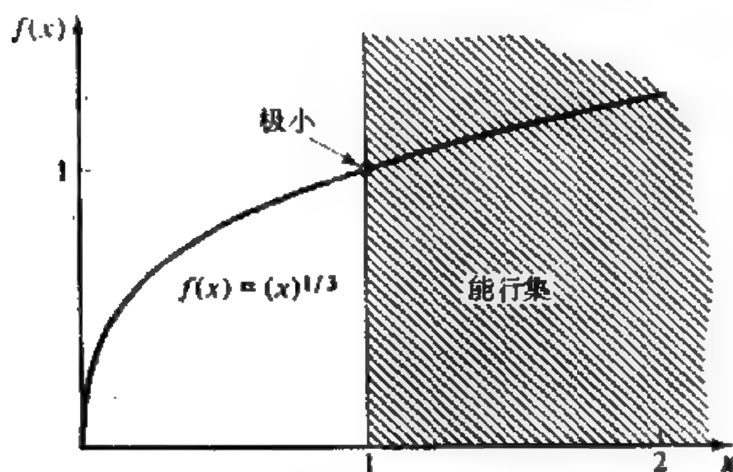
引入一个松弛变量 $\mu \geq 0_\phi$，得出

$$(x^{1/3} - w)^{1/3}[-]\mu = 0_\phi, \quad (7.129)$$

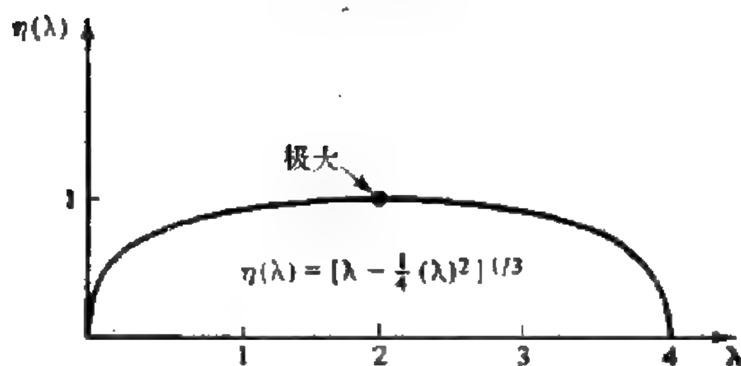
$$w = (x)^{1/3} - (\mu)^3. \quad (7.130)$$

因此，

$$\theta^*(\xi, \lambda) = \begin{cases} \sup_{x}\{\xi x + \lambda(x)^{1/3} - \lambda - x + 2\}^{1/3}, & (\lambda+1)^{1/3} \geq 0_\phi, \\ +\infty_\phi, & \text{其他} \end{cases} \quad (7.131)$$



(a)原有规划



(b)对偶规划

图 7.1 一对(h, ϕ)-凸规划

$$=\begin{cases}\left[\dfrac{(\lambda)^2}{4(1-\xi)}-\lambda+2\right]^{1/3}, & (\lambda+1)^{1/3}\geqslant 0,\ 和\ \xi\neq 1,\\ +\infty, & 其他.\end{cases}$$

(7.132)

而且，让 $\xi=0$，对偶规划为

$$\max[-]\theta^*(0_k,\ \lambda)=\eta(\lambda)=\left[\lambda-\frac{1}{4}(\lambda)^2\right]^{1/3} \tag{7.133}$$

受限制于

$$(\lambda+1)^{1/3}\geqslant 1, \tag{7.134}$$

它的最优解是 $\lambda^*=2$ 且 $\eta(\lambda^*)=1$. 原有规划和对偶规划在图 7.1 中说明. ▌

注意，这个例题中得到的对偶规划仅仅是利用种种扰动所能导出的许多可能的对偶规划中的一个. 虽然本例问题也是一个几何规划 (若我们还限制 x 是正的)，我们却没有把它作为 (log, log)-凸规划来处理. 因此可以看出，与一个确定的原有规划联系的一族对偶规划能够进一步扩大，其办法是找出不同的函数 h 和 ϕ，使一对原有-对偶规划变为 $(h,\ \phi)$-凸规划.

现在，我们来为几何规划求出一个对偶规划，相同于原来由 Duffin、Peterson、Zener 所推导的对偶规划. 我们不采用算术-几何平均值不等式[16] 或凸规划分析[19, 22, 23]，而使用 $(h,\ \phi)$-凸对偶性.

我们通过改写 (PGP) 着手. 令 y^k 是 $x\in R^n$ 的向量值函数，它的分量是 y_i，$i\in I_k$，满足

$$y_i=\prod_{j=1}^{n}(x_j)^{a_{ij}},\ i\in I_k,\ \ k=0,\ \cdots,\ m, \tag{7.135}$$

或者用 (log, log)-凸运算，有

$$y_i=(a_{i1}\odot x_1)\oplus\cdots\oplus(a_{in}\odot x_n), \tag{7.136}$$

于是得到

$$g_k(y^k)=\sum_{i\in I_k}c_i y_i,\quad k=0,\ \cdots,\ m, \tag{7.137}$$

规划 (PGP) 成为

(SGP) $$\min g_0(y^0) \tag{7.138}$$

受限制于

$$[-]g_k(y^k) \geqslant 0_{\log}, \quad k=1, \cdots, m, \tag{7.139}$$

这里 g_0 是(log, log)-凸函数, $[-]g_1, \cdots, [-]g_m$ 是(log, log)-凹函数, y_i 满足(7.136), 且

$$0_{\log} = (e)^0 = 1. \tag{7.140}$$

引入扰动向量 $u=(u^0, \cdots, u^m)^T$ 和 $w=(w_1, \cdots, w_m)^T$, 这里 u^k 是一向量, 其分量为 u_i, $i \in I_k$, 我们得到一个扰动问题:

$$\min g_0(y^0) = \sum_{i \in I_0} c_i y_i \tag{7.141}$$

受限制于

$$[-]g_k(y^k, w_k) = [-]\left(\sum_{i \in I_k} c_i y_i [+] w_k\right) \geqslant 0_{\log}, \quad k=1, \cdots, m, \tag{7.142}$$

$$y_i \ominus u_i = (a_{i1} \odot x_1) \oplus \cdots \oplus (a_{in} \odot x_n), \quad i \in I_k, \; k=0, \cdots, m. \tag{7.143}$$

注意, g_0 为(log, log)-凸函数, 这里第一个 log 对应于向量 y, g_k 分别关于 $(y^k, w_k)^T$ 是$[(h, v), \phi]$-凹函数, 且有 $h_i(y_i) = \log y_i$ 和 $v(w_k) = w_k$, 运算 $y_i \ominus u_i$ 在此处意味着 y_i/u_i, 因为我们选择了 $h_i(u_i) = \log u_i$.

这样, 原有规划(SGP)能重写为

(P_e) $$\min_{x, y} \theta(x, y, 0_{\log}, 0). \tag{7.144}$$

其中,

$$\theta(x, y, u, w) = g_0(y^0), \quad 若 \begin{cases} [-]g_k(y^k, w) \geqslant 0_{\log}, & k=1, \cdots, m, \\ y \ominus u = A \odot x, \end{cases} \tag{7.145}$$

这里第二个条件只是(7.143)的简略记法. 若(7.145)并不成立, 则 $\phi(x, y, u, w) = +\infty_{\log}$.

转到对偶规划之前, 我们计算一个后面要用到的共轭函数. 设 g 是 y 的一个(h, ϕ)-凸函数, 其中 $h(y) = (\log y_1, \cdots, \log y_n)^T$, $\phi(t) = \log t$. 假设 g 由下式给出:

$$g(y) = \sum_i c_i y_i, \tag{7.146}$$

这里 $c_i > 0$. 令 $\bar{h}(\xi) = \xi$, 并定义

$$g^*(\xi) = \sup_y \{(\xi^T y)_{I, \log, \log} [-] g(y)\}, \tag{7.147}$$

这里, 当 $\bar{h}(y) = y$ 时我们使用记号 $\bar{h} = I$, 即恒等函数.

要求读者证明:

$$g^*(\xi) = \begin{cases} \prod_i \left(\frac{\xi_i}{c_i}\right)^{\xi_i}, & \sum_i \xi_i = 1,\ \xi \geqslant 0, \\ +\infty_{\log}, & \text{其他}. \end{cases} \tag{7.148}$$

并且, g^* 是$(\bar{h}, \phi)$-凸函数, $\bar{h}$ 和 ϕ 与前相同.

现在计算 θ 的共轭函数. 定义

$$\begin{aligned} &\theta^*(\xi, \eta, \delta, \lambda) \\ &= \sup_{x, y, u, w} \{(\xi^T x)_{I, \log, \log} [+] (\eta^T y)_{I, \log, \log} [+] (\delta^T u)_{I, \log, \log} \\ &\quad [+] (\lambda^T w)_{v, \log} [-] \theta(x, y, u, w)\} \end{aligned} \tag{7.149}$$

$$\begin{aligned} &= \sup_{\substack{[-] g_k(y^k, w_k) \geqslant 0_{\log} \\ y \ominus u = A \odot x}} \{(\xi^T x)_{I, \log, \log} [+] (\eta^T y)_{I, \log, \log} \\ &\quad [+] (\delta^T u)_{I, \log, \log} [+] (\lambda^T w)_{I, \log} [-] \sum_{i \in I_0} c_i y_i\}. \end{aligned} \tag{7.150}$$

引入松弛变量 $\mu_k \geqslant 0_{\log}$, 我们得到

$$\begin{aligned} &\theta^*(\xi, \eta, \delta, \lambda) \\ &= \sup_{\substack{\mu_k \geqslant 0_{\log} \\ y \ominus u = A \odot x}} \{(\xi^T x)_{I, \log, \log} [+] (\eta^T y)_{I, \log, \log} \\ &\quad [+] (\delta^T u)_{I, \log, \log} [-] \sum_{k=1}^m \lambda_k [\cdot] (\sum_{i \in I_k} c_i y_i) \\ &\quad [-] (\lambda^T \mu)_{I, \log, \log} [-] \sum_{i \in I_0} c_i y_i\} \end{aligned} \tag{7.151}$$

$$= \begin{cases} \sup_{y \ominus u = A \odot x} \{(\xi^T x)_{I, \log, \log} [+] (\eta^T y)_{I, \log, \log} [+] (\delta^T u)_{I, \log, \log} \\ \quad [-] \sum_{k=1}^m \lambda_k [\cdot] (\sum_{i \in I_k} c_i y_i) [-] \sum_{i \in I_0} c_i y_i\}, & \lambda \geqslant 0 \\ +\infty_{\log}, & \text{其他}. \end{cases} \tag{7.152}$$

考虑 $\lambda \geqslant 0$ 的情况. 我们能写成

$$\theta^*(\xi, \eta, \delta, \lambda)$$

$$= \sup_{y\ominus u = A\odot x} \{(\xi^T x)_{I,\log,\log}[-](\delta^T y \ominus u)_{I,\log,\log}$$

$$[+]((\eta+\delta)^T y)_{I,\log,\log}[-]\sum_{k=1}^{m}\lambda_k[\cdot](\sum_{i\in I_k} c_i y_i)[-]\sum_{i\in I_0} c_i y_i\} \tag{7.153}$$

$$= \sup_x\{(\xi^T x)_{I,\log,\log}[-](\delta^T A\odot x)_{I,\log,\log}\}$$

$$[+]\sup_y\{((\eta+\delta)^T y)_{I,\log,\log}[-]\sum_{k=1}^{m}\lambda_k[\cdot](\sum_{i\in I_k} c_i y_i)$$

$$[-]\sum_{i\in I_0} c_i y_i\} \tag{7.154}$$

$$= S_1[+]S_2. \tag{7.155}$$

(7.155)中第一项是

$$S_1 = \sup_x\left\{\exp\left[\left(\xi_1 - \sum_{i=1}^{M} a_{i1}\delta_i\right)\log x_1 + \cdots + \left(\xi_n - \sum_{i=1}^{n} a_{in}\delta_i\right)\log x_n\right]\right\}, \tag{7.156}$$

同时，显然地

$$S_1 = \begin{cases} 0_{\log}, & \sum_{i=1}^{M} a_{ij}\delta_j = \xi_j,\ j=1, \cdots, n, \\ +\infty_{\log}, & 其他. \end{cases} \tag{7.157}$$

对于第二项，重新整理后我们得到

$$S_2 = \sup_{y_i, i\in I_0}\{((\eta^0+\delta^0)^T y^0)_{I,\log,\log}[-]\sum_{i\in I_0} c_i y_i\}$$

$$[+]\sup_{y_i, i\in I_1}\{((\eta^1+\delta^1)^T y^1)_{I,\log,\log}[-](\lambda_1[\cdot]\sum_{i\in I_1} c_i y_i)\}[+]\cdots$$

$$[+]\sup_{y_i, i\in I_m}\{((\eta^m+\delta^m)^T y^m)_{I,\log,\log}[-](\lambda_m[\cdot]\sum_{i\in I_m} c_i y_i)\}. \tag{7.158}$$

然后由(7.146)至(7.148)以及某些修改，我们有

$$S_2 = \begin{cases} \prod_{i\in I_0}\left(\frac{\eta_i+\delta_i}{c_i}\right)^{(\eta_i+\delta_i)}[+]\prod_{i\in I_1}\left(\frac{\eta_i+\delta_i}{c_i\lambda_1}\right)^{(\eta_i+\delta_i)}[+]\cdots \\ \quad [+]\prod_{i\in I_m}\left(\frac{\eta_i+\delta_i}{c_i\lambda_m}\right)^{(\eta_i+\delta_i)}, & \sum_{i\in I_k}(\eta_i+\delta_i) = \lambda_k,\ k=1, \cdots, m, \\ +\infty_{\log}, & 其他. \end{cases} \tag{7.159}$$

最后，有

$$\theta^*(0,\ 0,\ \delta,\ \lambda)=\prod_{i\in I_0}\left(\frac{\delta_i}{c_i}\right)^{\delta_i}\prod_{k=1}^{m}\prod_{i\in I_k}\left(\frac{\delta_i}{c_i\lambda_k}\right)^{\delta_i}, \tag{7.160}$$

此处，假设

$$\delta\geqslant 0, \tag{7.161}$$

$$\sum_{i\in I_0}\delta_i=1, \tag{7.162}$$

$$\sum_{k=0}^{m}\sum_{i\in I_k}a_{ij}\delta_i=0,\quad j=1,\ \cdots,\ n, \tag{7.163}$$

$$\sum_{i\in I_k}\delta_i=\lambda_k,\quad k=1,\ \cdots,\ m, \tag{7.164}$$

在其他情况，$\theta^*(0,\ 0,\ \delta,\ \lambda)=+\infty$，$(P_\theta)$的对偶规划由下式给出：

$$(D_{\theta^*})\qquad \max_{\delta,\lambda}[-]\theta^*(0,\ 0,\ \delta,\ \lambda)=\max_{\delta,\lambda}\Gamma(\delta,\ \lambda), \tag{7.165}$$

亦即

$$(\mathrm{DGP})\quad \max_{\delta,\lambda}\Gamma(\delta,\ \lambda)=\prod_{i\in I_0}\left(\frac{c_i}{\delta_i}\right)^{\delta_i}\prod_{k=1}^{m}\prod_{i\in I_k}\left(\frac{c_i}{\delta_i}\right)^{\delta_i}(\lambda_k)^{\lambda_k} \tag{7.166}$$

受限制于(7.161)～(7.164)．很清楚，这正是由 Duffin、Peterson 和 Zener[16] 及许多其他作者所得到的对偶规划．注意，Γ 关于 $h(t)=t$ 和 $\phi(t)=\log t$ 是$(h,\ \phi)$-凸函数．

这个对偶规划有许多有趣的特性．最重要的也许是，虽然原有问题具有非线性目标函数和非线性约束，然而约束 (7.161) 至 (7.164) 是线性的．按这个观点，几何规划属于这样一类非线性规划，即其对偶规划具有线性约束．这样重要的一类问题已由 Rockafellar[24] 研究过．这种规划的重要性在于，线性约束的非线性规划通常比非线性约束的规划更易于求解．然而，在几何规划情形，假如想通过求解对偶规划而得到原有规划的最优解，上述那种好处并非是显见的．当最优的对偶解已知时，去计算最优的原有变量并不总是方便的，特别，当某些原有约束在所寻求的最优点并非积极约束时更是如此．但是，至少出现一个有趣的情况，在这情况中求解对偶规划显然较为容易．当原有几何规划中单项式的总数 M 恰好比原有变量 n 多 1 时，对偶约束由非负变量的正方形线性方程组构成，而当这方程组线性无关时它可以有唯一解．我们

用一个例子来说明这一情况.

例 7.3.2

考虑几何规划

$$\min g_0(x)=\frac{(x_4)^2}{(x_1)^4(x_2)}+\frac{3(x_1)^2}{(x_2)^2} \tag{7.167}$$

受限制于 $x>0$ 和

$$g_1(x)=\frac{(x_2)(x_3)}{3}+\frac{3}{(x_1)^{1/2}(x_2)^{3/4}(x_3)}+\frac{9(x_2)^{1/2}}{2(x_3)(x_4)^{1/2}}\leqslant 1. \tag{7.168}$$

如前所述计算对偶规划,我们得到

$$\max \Gamma(\delta,\ \lambda)=\left(\frac{1}{\delta_1}\right)^{\delta_1}\left(\frac{3}{\delta_2}\right)^{\delta_2}\left(\frac{1}{3\delta_3}\right)^{\delta_3}\left(\frac{3}{\delta_4}\right)^{\delta_4}\left(\frac{9}{2\delta_5}\right)^{\delta_5}(\lambda_1)^{\lambda_1} \tag{7.169}$$

受限制于

$$\begin{aligned}
\delta_1+\delta_2 \qquad\qquad\qquad\qquad &=1,\\
-4\delta_1+2\delta_2 \qquad -\frac{1}{2}\delta_4 \qquad &=0,\\
-\delta_1-2\delta_2+\delta_3-\frac{3}{4}\delta_4+\frac{1}{2}\delta_5&=0,\\
\delta_3-\delta_4-\delta_5&=0,\\
2\delta_1 \qquad\qquad\qquad -\frac{1}{2}\delta_5&=0,
\end{aligned} \tag{7.170}$$

$$\delta_3+\delta_4+\delta_5=\lambda_1, \tag{7.171}$$

$$\delta\geqslant 0. \tag{7.172}$$

这个正方形的线性方程组(7.170)有唯一解

$$\delta_1^*=\frac{1}{4},\ \ \delta_2^*=\frac{3}{4},\ \ \delta_3^*=2,\ \ \delta_4^*=1,\ \ \delta_5^*=1.$$

因此 $\lambda_1^*=4$. 代入(7.169),我们得到 $\Gamma(\delta^*,\ \lambda^*)=384$. ∎

显然,在这个例题中,对于对偶规划,我们得到了唯一的能行解,因而也是最优解,因为在原有规划中我们有 $M=5$ 和 $n=4$. 若 $M>n+1$, 对偶规划通常存在无数能行解,为了寻找最优解,它的目标函数还必须求极大.在任何情况,最优原有解必须从最优对偶

解计算．在原有规划与对偶规划之间的关系能够从下面的对偶性定理中找到．结论将不加证明地给出，因为这里出现在一对原有–对偶规划中的函数是可变换为凸的，若我们利用这个事实，则有关结论便是第 5 章的一般凸对偶性理论的特殊情况． 等价地，可把第 5 章的结果推广到带有 (h, ϕ)–凸函数的非线性规划．

定理 7.13（弱对偶性定理）

若 x 满足 (PGP) 的约束，又 (δ, λ) 满足 (DGP) 的约束，则

$$g_0(x) \geqslant \Gamma(\delta, \lambda). \tag{7.173}$$

定理 7.14（强对偶性定理）

如果 (PGP) 是强相容的，又 x^* 是 (PGP) 的一个最优解，那末，

（i）规划 (DGP) 是相容的，并存在一个能行向量 (δ^*, λ^*)，使得

$$g_0(x^*) = \Gamma(\delta^*, \lambda^*). \tag{7.174}$$

（ii）存在一个 $\lambda^* \geqslant 0$，使得向量 (x^*, λ^*) 是下列 Lagrange 式对一切 $x > 0$ 和 $\lambda \geqslant 0$ 的鞍点：

$$L(x, \lambda) = g_0(x) [-] \sum_{k=1}^{m} \lambda_k [\cdot] g_k(x). \tag{7.175}$$

（iii）最优解 (δ^*, λ^*) 满足

$$\delta_i^* = \begin{cases} c_i \prod\limits_{j=1} \dfrac{(x_j^*)^{a_{ij}}}{g_0(x^*)}, & i \in I_0, \\ \lambda_k^* c_i \prod\limits_{j=1}^{n} (x_j^*)^{a_{ij}}, & i \in I_k,\ k = 1, \cdots, m. \end{cases} \tag{7.176}$$

（iv）若 (δ^*, λ^*) 是 (DGP) 的最优解，则在 $\lambda_k^* > 0$ 的条件下，

$$c_i \prod_{j=1}^{n} (x_j^*)^{a_{ij}} = \begin{cases} \delta_i^* \Gamma(\delta^*, \lambda^*), & i \in I_0, \\ \dfrac{\delta_i^*}{\lambda_k^*}, & i \in I_k,\ k = 1, \cdots, m. \end{cases} \tag{7.177}$$

注意，在 (7.175) 的方括号中的代数运算对应于 $\phi(t) = \log t$．

希望证明这个定理的读者可以从定理 5.16（或它的等价结果）出发，那里断定强相容性包含稳定性． 一旦稳定性建立了，就可由第 5 章并通过计算次梯度得出结论的其余部分．最后两个定理首先由 Duffin、Peterson、Zener[16] 证明，在那里还可找到其他对

偶性结果.

练　习

7.A. 给定一个实向量 $q\in R^p$ 和实 $p\times p$ 矩阵 M，假设要寻求满足下列条件的向量 w 和 z：

$$w=q+Mz,\ w\geqslant 0,\ z\geqslant 0, \tag{7.178}$$

$$z^Tw=0. \tag{7.179}$$

这个问题被 Cottle 和 Dantzig[9] 称为**互补性基本问题**. 试证明，对非负变量 $x\geqslant 0$ 求解由(7.1)和(7.2)给出的凸二次规划问题，等价于求解上面介绍的一个相应的互补性问题.

[提示：定义

$$u=c+Qx-A\lambda,\quad v=A^Tx-b. \tag{7.180}$$
]

7.B. 求由(7.20)至(7.22)给出的规划(DQP)的对偶规划].

7.C. 试证(7.70)至(7.75)是带可分离偿付的随机线性规划问题(CC)的条件.

7.D. 试证(PLRP)的对偶规划(DLRP)的确由(7.80)至(7.83)给出.

7.E. 对于带有线性偿付函数(7.77)的随机规划，构成其存在性、特性及对偶性定理.

7.F. 设 g 是正项式，证明 $1/g$ 是(log, log)-凹函数.

7.G. 证明(7.148)，并证明 g^* 是$(\bar{h},\phi)$-凸的.

7.H. 验证关系(7.156)至(7.166).

7.I. 从最优对偶解来计算 x^*，继续分析例 7.3.2. 验证在这种情况下(7.174)成立.

7.J. 指出第 5 章和第 6 章中哪些结果对于定理 7.14 的证明是必需的.

7.K. 假设有下列线性的极小极大规划

(PM)
$$\min_x\left\{\max_{1\leqslant i\leqslant p}\left(\sum_{j=1}^n c_{ij}x_j+d_i\right)\right\} \tag{7.181}$$

受限制于

$$\sum_{j=1}^n a_{kj}x_j>b_k,\quad k=1,\cdots,m. \tag{7.182}$$

试证对应于(PM)的凸对偶规划是

(DM)
$$\max\sum_{i=1}^p d_i\eta_i+\sum_{k=1}^m b_k\lambda_k \tag{7.183}$$

受限制于

$$\sum_{i=1}^p \eta_i=1, \tag{7.184}$$

$$\sum_{k=1}^{m} a_{kj}\lambda_k = \sum_{i=1}^{p} c_{ij}\eta_i, \quad j=1, \cdots, n, \tag{7.185}$$

$$\eta \geqslant 0, \quad \lambda \geqslant 0. \tag{7.186}$$

7.L. 考虑代数规划[22]:

$$\min G_0(x) \tag{7.187}$$

受限制于

$$G_k(x) \leqslant 1, \quad k=1, \cdots, m, \tag{7.188}$$

这里，对于 $k=0, 1, \cdots, m$,

$$G_k(x) = \Big(\sum_{i \in I_k} \{w_i [\sum_{l \in L_i} (P_l(x))^{\alpha_l r_k} + (R_l(x))^{\beta_l r_k}]^{1/r_k}\}^{p_k}\Big)^{1/p_k}, \tag{7.189}$$

其中 P_l、R_l 是 $x>0$ 的正项式，α_l、β_l、p_k、r_k、w_i 是正的常数．找出它的一个对偶规划．

7.M. 若让上题中 $r_k \to +\infty$，得到

$$G_k(x) = \Big(\sum_{i \in I_k} \{w_i \sum_{l \in L_i} \max[(P_l(x))^{\alpha_l}, (R_l(x))^{\beta_l}]\}^{p_k}\Big)^{1/p_k}. \tag{7.190}$$

在这个情况下修改得到的对偶是什么？构成并求解一个代数规划的数值例子，其中函数 G_k 由(7.190)给出，求解时，如同例 7.3.2 那样，利用它的对偶．更一般的包含广义正多项式的代数规划已由 Avriel 和 Gurovich[9] 和 Gurovich[18] 处理．

参 考 文 献

1. AVRIEL, M., M. J. RIJCKAERT, and D. J. WILDE, (Eds.), *Optimization and Design*, Prentice-Hall, Englewood Cliffs, N.J., 1973.
2. AVRIEL, M., and V. GUROVICH, "A Condensation Algorithm for a Class of Algebraic Programs," *Operations Research*, to appear.
3. AVRIEL, M., and A. C. WILLIAMS, "Stochastic Linear Programming with Separable Recourse Functions," Mobil R & D Corp. Central Research Div. Progress Memorandum, Princeton, N.J. March 1967.
4. AVRIEL, M., and A. C. WILLIAMS, "Stochastic Programming with Separable Recourse Functions," paper presented at the 33rd ORSA National Meeting, San Francisco, May 1968.
5. BEALE, E. M. L., "Numerical Methods," in *Nonlinear Programming*, J. Abadie (Ed.), North-Holland Publishing Co., Amsterdam, 1967.
6. BOOT, J. G. C., *Quadratic Programming*, North-Holland Publishing Co., Amsterdam, 1964.
7. COTTLE, R. W., "Symmetric Dual Programs," *Quart. Appl. Math.*, **21**, 237–243 (1963).

8. COTTLE, R. W., "The Principal Pivoting Method of Quadratic Programming," in *Mathematics of the Decision Sciences, Part I*, G. B. Dantzig and A. F. Veinott (Eds.), American Mathematical Society, Providence, R.I., 1968.

9. COTTLE, R. W., and G. B. DANTZIG, "Complementary Pivot Theory of Mathematical Programming," *Linear Alg. Appl.*, **1**, 103–125 (1968).

10. COTTLE, R. W., and W. C. MYLANDER, "Ritter's Cutting Plane Method for Nonconvex Quadratic Programming," in *Integer and Nonlinear Programming*, J. Abadie (Ed.), North-Holland Publishing Co., Amsterdam, 1970.

11. DANTZIG, G. B., *Linear Programming and Extensions*, Princeton University Press, Princeton, N.J., 1963.

12. DENNIS, J. B., *Mathematical Programming and Electrical Networks*, Technology Press, Cambridge, Mass., 1959.

13. DORN, W. S., "Duality in Quadratic Programming," *Qudrt. Appl. Math.*, **18**, 155–162 (1960).

14. DORN, W. S., "A Duality Theorem for Convex Programs," *IBM J. Res. Dev.*, **4**, 407–413 (1960).

15. DORN, W. S., "Self-dual Quadratic Programs," *J. SIAM*, **9**, 51–54 (1961).

16. DUFFIN, R. J., E. L. PETERSON, and C. ZENER, *Geometric Programming—Theory and Applications*, John Wiley & Sons, New York, 1967.

17. Fenchel, W., "On Conjugate Convex Functions," *Can. J. Math.*, **1**, 73–77 (1949).

18. GUROVICH, V., "Extensions of Signomial Programming: Numerical Solutions," M.Sc. thesis, Technion, Israel Institute of Technology, Haifa, 1974.

19. HAMALA, M., "Geometric Programming in Terms of Conjugate Functions," Center for Operations Research and Econometrics, Université Catholique de Louvain Discussion Paper No. 6811, Louvain, Belgium, June 1968.

20. HARDY, G. H., J. E. LITTLEWOOD, and G. PÓLYA, *Inequalities*, 2nd ed., University Press, Cambridge, England, 1952.

21. KELLER, E. L., "The General Quadratic Optimization Problem," *Math. Prog.*, **5**, 311–337 (1973).

22. PETERSON, E. L., "Geometric Programming and Some of Its Extensions," in *Optimization and Design*, M. Avriel, M. J. Rijckaert, and D. J. Wilde (Eds.), Prentice-Hall, Englewood Cliffs, N.J., 1973.

23. ROCKAFELLAR, R. T., *Convex Analysis*, Princeton University Press, Princeton, N.J., 1970.

24. ROCKAFELLAR, R. T., "Some Convex Programs Whose Duals are Linearly Constrained," in *Nonlinear Programming*, J. B. Rosen, O. L. Mangasarian, and K. Ritter (Eds.), Academic Press, New York, 1970.

25. SENGUPTA, J. K., *Stochastic Programming—Methods and Applications*, North-Holland Publishing Co., Amsterdam, 1972.

26. Vajda, S., *Probabilistic Programming*, Academic Press, London, 1972.

27. Walkup, D. W., and R. J. B. Wets, "Stochastic Programs with Recourse," *SIAM J. Appl. Math.*, **15**, 1299–1314 (1967).

28. Wets, R. J. B., "Stochastic Programs with Recourse: A Survey I," Boeing Scientific Research Laboratories Document D1-82-0882, Seattle, Wash. August 1969.

29. Williams, A. C., "On Stochastic Linear Programming," *SIAM J. Appl. Math.*, **13**, 927–940 (1965).

30. Williams, A. C., "Approximation Formulas for Stochastic Linear Programming," *SIAM J. Appl. Math.*, **14**, 668–677 (1966).

31. Zang, I., "Generalized Convex Programming," D.Sc. dissertation, Technion, Israel Institute of Technology, Haifa, 1974.

32. Zener, C., *Engineering Design by Geometric Programming*, John Wiley & Sons, New York, 1971.

33. Ziemba, W. T., "Computational Algorithms for Convex Stochastic Programs with Simple Recourse," *Operations Research*, 18, 414–431 (1970).